



Research Note 2017-01

Identifying and Validating Selection Tools for Predicting Officer Performance and Retention

Teresa L. Russell, Editor

Cheryl J. Paullin, Editor

Human Resources Research Organization

Peter J. Legree, Editor

Robert N. Kilcullen, Editor

Mark C. Young, Editor

U.S. Army Research Institute

May 2017

**United States Army Research Institute
for the Behavioral and Social Sciences**

Approved for public release, distribution unlimited.

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

**Department of the Army
Deputy Chief of Staff, G1**

Authorized and approved:

**MICHELLE SAMS, Ph.D.
Director**

Research accomplished under contract
for the Department of the Army by

Human Resources Research Organization

Technical review by

Rebekkah Beeco, U.S. Army Research Institute

NOTICES

DISTRIBUTION: This Research Note has been submitted to the Defense Technical Information Center (DTIC). Address correspondence concerning reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: DAPE-ARI-ZX, 6000 6th Street (Bldg. 1464 / Mail Stop: 5610), Fort Belvoir, Virginia 22060-5610.

FINAL DISPOSITION: Destroy this Research Note when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Note are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE					
1. REPORT DATE (dd-mm-yy) May 2017		2. REPORT TYPE Final		3. DATES COVERED (from....to) December 2011 to December 2014	
4. TITLE AND SUBTITLE Identifying and Validating Selection Tools for Predicting Officer Performance and Retention				5a. CONTRACT OR GRANT NUMBER W5J9CQ-12-C-0004	
				5b. PROGRAM ELEMENT NUMBER 622785	
6. EDITORS Teresa L. Russell, Cheryl J. Paullin; Peter J. Legree, Robert N. Kilcullen, & Mark C. Young				5c. PROJECT NUMBER A790	
				5d. TASK NUMBER	
				5e. WORK UNIT NUMBER 311	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Human Resources Research Organization 66 Canal Center Plaza, Suite 700 Alexandria, Virginia 22314				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg 1464/Mail Stop 5610) Fort Belvoir, VA 22060-5610				10. MONITOR ACRONYM ARI	
				11. MONITOR REPORT NUMBER Research Note 2017-01	
12. DISTRIBUTION/AVAILABILITY STATEMENT: Distribution Statement A. Approved for public release, distribution unlimited.					
13. SUPPLEMENTARY NOTES ARI Research POCs: Dr. Peter J. Legree, and Dr. Mark C. Young, Personnel Assessment Research Unit					
14. ABSTRACT (<i>Maximum 200 words</i>): The U.S. Army must commission officers who are likely to perform well as junior officers, fit into the Army's culture, demonstrate leadership potential for higher ranks, and be motivated to stay beyond their initial Active Duty Service Obligation (ADSO). To address this requirement, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has undertaken a program of research to enhance officer selection, assignment, and retention. The primary purpose of this effort was to evaluate a number of predictor measures against officer performance and career intentions. The core activity was a concurrent, criterion-related validation project in which several predictor and criterion measures were administered to over 800 early- and mid-career officers. Results showed that a number of measures were useful predictors of officer performance and retention. The Rational Biodata Inventory (RBI), the Leader Knowledge Test (LKT), the Objective-Format Consequences Test, and College GPA were good predictors of technical, managerial, leadership, and effort/discipline performance dimensions.					
15. SUBJECT TERMS Officer selection, predictor measures, criterion measures					
SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT Unlimited Unclassified	20. NUMBER OF PAGES 84	21. RESPONSIBLE PERSON (Name and Telephone Number) Tonia S. Heffner (703) 545-4408
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified			

Research Note 2017-01

**Identifying and Validating Selection Tools for
Predicting Officer Performance and Retention**

Teresa L. Russell, Editor

Cheryl J. Paullin, Editor

Human Resources Research Organization

Peter J. Legree, Editor

Robert N. Kilcullen, Editor

Mark C. Young, Editor

Personnel Assessment Research Unit

Tonia S. Heffner, Chief

May 2017

Approved for public release, distribution unlimited.

IDENTIFYING AND VALIDATING SELECTION TOOLS FOR PREDICTING OFFICER PERFORMANCE AND RETENTION

EXECUTIVE SUMMARY

Research Requirement:

The U.S. Army is undergoing a major transition in its personnel structure as the Army's end strength is projected to drop significantly through 2017. These changes will increase the challenges and criticality of selecting those officers who have the most promising potential to lead our future, rapidly changing Army. In this environment, the Army needs officers who will perform well as junior officers, excel at higher leadership levels, and be motivated to pursue a career of long-term service. With those changes in mind, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) initiated a program of research on methods of enhancing selection, assignment, and retention of officers.

The current effort builds upon earlier ARI research that (a) identified the major duty requirements across a range of company and field-grade officer leadership positions (i.e., platoon leader [PL], company commander [CC], battalion staff officers [Bn XO, Bn S3], and battalion commander [Bn Cdr]), and (b) developed both predictor and criterion measures relevant to Army officer performance and continuance. Using the knowledge and lessons learned from those earlier efforts, the current effort "Identifying and Validating Selection Tools for Officer Performance and Retention" was conducted to further inform the development and implementation of the Army's officer selection measures and procedures, while enhancing the development and understanding of relevant scientific theory.

The primary purpose of the current project was to evaluate a number of predictor measures against officer performance and career intentions. The core activity was a concurrent, criterion-related validation project in which we administered several predictor and criterion measures to early- and mid-career officers. As officers move from company grade to field grade leadership positions (e.g., battalion command and staff positions), their numbers greatly diminish, making it difficult to conduct criterion-related validations of new predictors. Accordingly, we applied a synthetic validation strategy for evaluating the value of our predictor measures for these higher-level positions.

Procedure:

Officers completed a computer-administered test battery composed of cognitive and non-cognitive predictors, a self-report career history instrument, and job knowledge measures. According to their duty position, officers completed different combinations of predictor and criterion instruments. Their supervisors provided performance ratings. Data from administrative records were collected for participants. Of the 877 officers who participated in the data collections, 836 provided data that were useable for at least some portion of the instruments.

The predictor battery included a wide array of measures. Cognitive measures included the Figural Reasoning Test and Objective-Format Consequences Test. Non-cognitive measures

included the Rational Biodata Inventory (RBI), Modified Career Path Appreciation (MCPA) test, Leader Knowledge Test (LKT), Work Values, and Virtual Leadership Judgment Simulation (VLJS).

Numerous criterion variables (e.g., supervisor ratings of performance, archival records, and job knowledge test [JKT] scores) were combined to form four broad performance dimensions: (a) Technical Task Proficiency, (b) Management, Administration, and Communication Proficiency, (c) Physical Fitness and Effort, and (d) Leadership, Supervision, and Personal Discipline. The four performance dimensions were combined into an Overall Performance composite.

Findings:

Simple bivariate correlations indicated that the RBI Fitness Motivation scale was the strongest predictor of Overall Performance for all positions. Several other RBI scales also predicted Overall Performance significantly in all positions, namely Achievement Orientation, Goal Expectations, Peer Leadership, and Tolerance for Ambiguity. The LKT, which was not administered to battalion XO or S3s, was significantly related to overall performance for both PLs and CCs. RBI Written Communication and RBI Hostility to Authority were significantly related to Overall Performance for battalion staff, but not for PL and CC positions. RBI Tolerance for Injury and College GPA were both significant predictors of Overall Performance in PL and CC positions, but not for battalion staff positions.

Platoon leader and company commander positions had sample sizes sufficient for use of more advanced statistics. Using Bayesian Model Averaging (BMA), a regression model was estimated for every possible combination of predictors. The results of the BMA showed that the strongest predictors of the four performance dimensions for both PL and CC positions were the Consequences Test and two RBI scales: Fitness Motivation and Tolerance for Injury. For CCs, a consistently strong predictor of performance was self-reported GPA. In addition, the RBI scales of Army Identification and Goal Expectations were reasonable predictors of Career Intentions for both positions. Comparable results were obtained using conventional regression approaches.

A synthetic validation approach was used for positions for which the sample sizes were too small to use regression-based approaches (i.e., Bn XO, S3, and Cdr). For the battalion command and staff positions, synthetic validity estimates suggested that the predictor battery is likely a valid predictor of performance. For the positions with larger sample sizes (i.e., PL and CC), the synthetic validity estimates were comparable to those obtained using the BMA approach.

Utilization and Dissemination of Findings:

The research findings have implications for selection of Officer Candidate School (OCS) candidates, evaluation of Reserve Officer Training Corps (ROTC) cadets, and selection of ROTC 4-year scholarship recipients. The results will be used to identify measures or scales that, if added to current selection processes, would help identify candidates or cadets who are likely to perform well in officer positions and stay in the Army beyond their active duty service obligation.

IDENTIFYING AND VALIDATING SELECTION TOOLS FOR PREDICTING OFFICER PERFORMANCE AND RETENTION

CONTENTS

	Page
Chapter 1: Meeting the Army's Needs for Committed, High-Performing Officers	1
ARI's Officer Research Program.....	1
The Officer Job Analysis Project.....	2
Predictors and Criteria for Officer Selection	3
Objectives of the Current Project	3
Chapter 2: Concurrent Validation Data Collection.....	5
Preparations for Data Collection	5
On-Site Data Collection Procedures	6
Database Development	7
Database Screening and Cleaning	8
Sample Demographics	8
Chapter 3: Measurement of the Criterion Domain	13
Criterion Measures.....	13
Performance Rating Scales	13
Objectively Scored Job Knowledge Tests (JKTs)	14
Self-Report: Career History Survey (CHS)	16
Archival Variables	16
Development of the Career Intentions Score.....	18
Development of Performance Dimension Scores	18
Psychometric Properties of Criterion Scores	20
Development of the Overall Performance Composite.....	24
Chapter 4: Predictor Measures and Psychometrics.....	27
Predictor Measures	27
Figural Reasoning Test.....	28
Rational Biodata Inventory (RBI)	28
Modified Career Path Appreciation (MCPA)-Phrases	28
Leader Knowledge Test (LKT)	29
Work Values	30
Objective-Format Consequences Test.....	30
Virtual Leadership Judgment Simulation (VLJS)	30
College GPA Category	31
Archival Predictor Variables	31
Psychometric Properties of Predictor Scores.....	31

CONTENTS (Continued)

	Page
Chapter 5: Criterion-Related Validation Analyses and Results.....	36
Analysis Approach.....	36
Bivariate Correlations	36
Regression Analyses	40
Summary	46
Synthetic Equations	47
Choosing Predictors for Each Performance Dimension	47
Building and Evaluating Synthetic Equations	48
Summary.....	50
Chapter 6: Summary and Conclusions.....	51
Objectives	51
Summary of Validation Approach and Key Findings	52
References	56

APPENDICES

Appendix A. Supplemental Criterion Psychometric Tables	A-1
Appendix B. Supplemental Predictor Psychometric Tables	B-1
Appendix C. Supplemental Validity Tables	C-1

CONTENTS (Continued)

	Page
TABLES	
Table 2.1.	Instruments Administered by Position..... 6
Table 2.2.	Number of Participants by Site 6
Table 2.3.	Final Sample by Army Component, Pay Grade, Commission Source, Level of Education, and Position 9
Table 2.4.	Final Sample by Branch and Position 10
Table 2.5.	Final Sample by Gender, Ethnicity, Race, Marital Status, and Position 11
Table 2.6.	Time in Service and Time in Grade for the Final Sample 12
Table 3.1.	Performance Rating Scale (PRS) Items 14
Table 3.2.	Preliminary Performance Components 19
Table 3.3.	Confirmatory Factor Analyses (CFA) Results..... 20
Table 3.4.	Final Criterion Dimensions and Measures..... 22
Table 3.5.	Criterion Score Descriptive Statistics 24
Table 3.6.	Criterion Score Intercorrelations..... 24
Table 3.7.	Criterion Dimension Subgroup Differences 24
Table 3.8.	Performance Dimension Mean Ratings and Weights 25
Table 4.1.	Predictor Measurement Plan 27
Table 4.2.	Rational Biodata Inventory (RBI) Scale Definitions 29
Table 4.3.	Predictor Score Descriptive Statistics 32
Table 4.4.	Predictor Score Subgroup Differences..... 34
Table 5.1.	Bivariate Validities for Platoon Leader 37
Table 5.2.	Bivariate Validities for Company Commander 38
Table 5.3.	Bivariate Validities for Battalion Staff Positions..... 39
Table 5.4.	Predictor Measures Included in Regression Analyses 40
Table 5.5.	Predictor Composite Results for Platoon Leaders 43
Table 5.6.	Predictor Composite Results for Company Commanders 44
Table 5.7.	Validity Evidence of Optimized Predictor Composites..... 45
Table 5.8.	Synthetic Predictor Weighting Scheme 48
Table 5.9.	Synthetic Predictor Weights by Performance Dimension 49
Table 5.10.	Comparison of Synthetic and BMA Weights by Position 50
Table 5.11.	Comparison of Synthetic and BMA Validity Coefficients by Position..... 50

IDENTIFYING AND VALIDATING SELECTION TOOLS FOR PREDICTING OFFICER PERFORMANCE AND RETENTION

CHAPTER 1: MEETING THE ARMY'S NEEDS FOR COMMITTED, HIGH-PERFORMING OFFICERS

Robert Kilcullen and Roy Campbell

The Army personnel system is quickly evolving, and large changes are expected for its training, procurement, promotion, and retention systems. Significant changes are occurring across the spectrum of the Army. An Army of over 565,000 as recently as 2011 may see end strengths of approximately 460,000 by 2017 (Tan, 2016). The Army may experience the necessary culling of officers as well as imbalances of officers across branches. There may also be changes in officer promotion and retention policies, officer procurement levels, and officer selection and qualification standards.

The deployment environment will likely change as well. An Army honed on 15 years of operational combat deployments may be refocused on the prevention of war. As fluid and undefined as the opposing forces were in Iraq and Afghanistan, the future may be one where sources of threat change rapidly among opportunists that are neither clearly friend nor enemy (Department of the Army, 2013).

To meet these emerging challenges with a smaller force, it will be important for the Army to be able to select high quality officers who are motivated to remain in the Army. At times, producing the right quantity of officers has received higher priority. Given the emerging demands of officer capabilities, giving precedence to quantity over quality may not be a viable alternative. Some commentators argue that the Army must apply a set of rigorous and uniform standards that incorporate high learning achievement, strong aptitude levels, problem solving skills, mental acuity, and intellectual agility (Wardzynski, Lyle, & Colarusso, 2009).

Unlike enlisted accessioning, the officer pipeline must start long before the individual is needed. The Reserve Officer Training Corps (ROTC) normally has a 4-year lead time to commissioning. The United States Military Academy (USMA) also requires a full four years to produce a commissioned officer. Officer Candidate School (OCS) provides the quickest source for officer accession, but even that process can be relatively lengthy; significantly exceeding the requirements of the 12-week OCS course itself. Changes to officer accession policy applied today may not be fully felt until the early to mid-2020s. Officer procurement programs must therefore take a long-term view based on studied projections about future conditions and operating climate.

ARI's Officer Research Program

The Army must commission officers who are likely to perform well as junior officers, fit into the Army's culture, demonstrate leadership potential for higher ranks, and be motivated to stay beyond their initial Active Duty Service Obligation (ADSO). To address this requirement, ARI has undertaken a program of research to enhance officer selection, assignment, and retention. A

secondary objective of this program is to advance various areas of scientific research by identifying opportunities to test scientific theories that have implications for the Army's ability to select, assign, develop, and retain effective leaders.

The first project in this series was the officer job analysis (Paullin et al., 2014; Paullin et al., 2011). It resulted in descriptions of major duties required in platoon leader, company commander, battalion operations officer (S3), battalion executive officer (XO), and battalion commander positions as well as branch-specific major duties (MDs) for junior infantry and armor officers. Predictor and criterion development drew heavily on the job analysis as well as ARI's ongoing research efforts in the ROTC program (Bynum & Legree, 2015; Legree, Kilcullen, Putka & Wasko, 2014; Bynum & Russell, 2013; Putka, 2009) and the OCS program (Allen, Bynum, Erk, Babin, & Young, 2012; Allen & Young, 2012; Russell & Tremble, 2011). The third project, Identifying and Validating Selection Tools for Predicting Officer performance and Retention, the subject of this report, examined the validity of the selection measures for predicting various job performance criteria.

The Officer Job Analysis Project

The purpose of the job analysis was to identify Army officer performance requirements for use in officer selection and assignment. The project focused on requirements for five positions: platoon leader, company commander, battalion XO, battalion S3, and battalion commander. In addition, the project focused on Army-wide requirements, supplemented by an effort to identify branch-specific requirements for several branches.

Performance requirements are the duties that officers are expected to be able to do on the job with a reasonable degree of proficiency. Paullin et al. (2011) compiled lists of MDs performed by all officers in each of the five positions listed above based on Army doctrine and training materials. They also (a) explored an approach for identifying branch-specific technical MDs in five branches: Armor, Infantry, Signal, Quartermaster, and Transportation, and (b) compiled lists of skills, abilities, and other characteristics (SAOs) required to perform the MDs based on analyses of Army leader competencies and supplemented by research on leadership and turnover/retention in other branches of the U.S. military and in the civilian sector. Finally, Paullin et al. (2011) asked Army officers to review the lists to ensure they reflected current Army practices. Qualitative feedback was used to revise the lists of MDs and SAOs, and, in turn, quantitative evaluations of importance and frequency of performance from 180 Army officers, most at the captain level were collected.

Paullin et al. (2011) identified a number of important MDs and SAOs for each of the five positions. The lists of MDs and SAOs were vetted by officers at each level, although there were very few reviewers at the higher-ranking levels. The authors also prepared draft descriptions of branch-specific MDs and obtained a reasonable level of review and consensus for the Armor and Infantry branches. For the other branches, Paullin et al. (2011) discovered that there is considerable variability in technical duty assignments within the branch, depending on the officer's functional duty assignment. The research team also found that the Army's ongoing transition to a modular, brigade-centric structure led to significant changes in branch-specific duties. The changing duty requirements had not been fully documented in published Army

doctrine. As a consequence, the approach that worked for identifying leadership and Army-wide technical MDs did not work very well for identifying branch-specific technical MDs.

Predictors and Criteria for Officer Selection

In the second project, Paullin et al. (2012) used the results of the job analysis to define the predictor and criterion domains. A technical advisory committee (TAC) worked with the project team to develop predictor and criterion measurement plans. Using the MDs from the job analysis as the foundation for the criterion domain, Paullin et al. planned four types of criterion measures: archival variables, performance ratings, job knowledge tests (JKTs), and self-reports of attitudes, career history, and accomplishments. Most of the measures were intended to assess aspects of performance common to all officers, such as leadership and Army-wide technical duties. They also developed measures of branch-specific technical duties at the platoon leader and company commander levels for two officer branches: Armor and Infantry. The latter were intended as explorations into developing criterion measures that could be used to evaluate officer branch assignments (classification). With the SAOs from the job analysis defining the predictor domain, Paullin et al. identified predictor measures for inclusion in future validation work as well as three measures needing some pilot work in the current project: the Consequences Test, the Goal Orientation scale, and the Modified Career Path Appreciation (MCPA) Phrases assessment.

Draft predictor and criterion measures were pilot tested. Paullin et al. (2012) obtained data from applicant-like samples including ROTC cadets attending the Leadership Development Assessment Course (LDAC) and OCS candidates. Criterion measures were pilot-tested on active duty officers at several training schools. The pilot test data were used to refine and shorten both predictor and criterion measures for use in the follow-on criterion-related validation project.

Objectives of the Current Project

The primary purpose of the current project was to evaluate the validity of a number of predictor measures against officer performance and career intentions. The core activity was a concurrent, criterion-related validation project in which we administered several promising predictor measures and the newly-developed criterion measures to early- and mid-career officers. A Technical Advisory Committee composed of Dr. Ken Pearlman, Dr. Trueman Tremble, Dr. Deirdre Knapp, and Mr. Roy Campbell provided advice and guidance, particularly on predictor and criterion selection early in the project.

Several new measures were developed as a part of this effort, including (a) a rating-based standardized scoring approach for the Consequences Test, (b) an animated simulation measure of leadership judgment that can be administered and scored via computer, (c) an essay test and essay scoring process to measure written communication skills, and (d) criterion measures that focused on branch-specific technical duties for three officer branches: Signal, Transportation, and Quartermaster. Several data collections supported instrument development. For example, data were collected using Mechanical Turk for the purpose of developing the objective-format Consequences Test. The essay test questions were administered to ROTC cadets during LDAC. Professors of Military Science provided input and pilot test data for the leadership judgment

simulation, and Army Training and Doctrine Command (TRADOC) posts supported the development of the branch-specific criterion measures.

Our general approach to the concurrent validation project was guided by three main principles. First, we wanted evaluate predictors and criterion measures that ARI had previously developed. This research used many of the most promising measures used in prior OCS and ROTC projects. Second, to measure job performance broadly, we gathered a wide array of criterion variables and went through a number of steps to integrate them into a small set of meaningful scores. Third, we wanted to evaluate the validity of the measures for predicting performance in higher ranking positions. However, we knew that sample sizes at the higher ranks would be small and insufficient to support regression analyses. To estimate the validities for the higher ranks, we followed a synthetic validation approach and created synthetic equations.

Over 800 mid-career officers participated in the validation project. Officers completed a computer-administered test battery composed of cognitive and non-cognitive predictors, a self-report career history instrument, and job knowledge measures. Their supervisors provided performance ratings. At the platoon leader, company commander and battalion staff levels, sample sizes were sufficient for examination of the empirical relationships between predictor and criterion scores. At the battalion command level, we collected descriptive and qualitative data from a smaller sample of incumbent officers. We computed synthetic equations to estimate the validity of the predictor battery for battalion commanders and also for battalion staff officers.

This report documents the procedures and results of the criterion-related validation project. Chapter 2 describes the concurrent validation data collection. Chapter 3 describes the measurement of the criterion domain. Chapter 4 describes the predictor measures and their psychometric properties. Chapter 5 describes validation analyses and results. Chapter 6 discusses the implications of these findings for Army officer selection, assignment and retention.

CHAPTER 2: CONCURRENT VALIDATION DATA COLLECTION

Charlotte Campbell, Laura Ford, Bethany Bynum, Jeff Barnes, and William Holden

Eight hundred and seventy-seven U.S. Army active duty officers from U.S. Army Forces Command (FORSCOM) units participated in the data collection. Data collection teams administered computer-based predictor and criterion measures to battalions tasked for participation at 11 data collection sites. Archival data from the Officer Longitudinal Research Database (OLRDB) and ROTC's Leadership Development and Assessment Course (LDAC) were merged with the predictor and criterion data to create the final database. The final sample size was 836 after data quality and screening checks. This chapter describes data collection and database development procedures and the demographic characteristics of the final sample.

Preparations for Data Collection

Preparation for the concurrent validation data collection involved decisions regarding: where data collections would be conducted; how officers would be tasked and scheduled; how the assessments would be administered; and how data collectors would be trained and resourced. The considerations underlying those decisions were not independent.

We collected data at FORSCOM locations where we could access intact units, such that officers and their supervisors could participate at the same time, thus facilitating the collection of supervisor ratings for junior officers in the sample. We also collected data at one TRADOC locations where officers in training would not have a supervisor who had directly and recently observed his/her job performance. We utilized the "Umbrella Week System" to collect data from Army posts and individual units. While sample selection was not entirely random, we were able to access a broad cross-section of FORSCOM locations.

Each RSR requested 12 platoon leaders (PLs), 4 company commanders (CCs), and the battalion commander, XO, and S3 from each of five or six battalions at each location. The schedule proposed to the point of contact (POC) always specified that we would do one battalion per day, with PLs participating in a morning session, then CCs in the early afternoon, and battalion commanders, XOs, and S3s last. This design allowed us to (a) obtain Informed Consents from lower echelon officers without any appearance of coercion from their supervisors, (b) limit supervisor ratings to only those of their lower echelon leaders who had participated, and (c) set up the data collection equipment at a location convenient for each battalion.

Table 2.1 shows, by title, the array of instruments for each officer position. For PLs and CCs, we constructed three administration sequences to minimize fatigue effects. For the last six data collections, we also administered the leadership judgment simulation (i.e., the Virtual Role-Play [VLJS]) to PLs. Once the procedures were documented, a Data Collection Manual was prepared and data collectors were trained. The Manual and the training were comprehensive, covering coordination with POCs, arrangements for travel and for shipping the laptops, scheduling the different leader groups, procedures for sign-ins and starting the assessments, responding to unexpected disruptions to the schedule and arrangements, downloading data, and conducting follow-ups to obtain ratings.

On-Site Data Collection Procedures

In all, we made eleven data collection trips, as shown in Table 2.2. While most sites were FORSCOM locations, we administered a subset of the predictors and criterion instruments to 27 O5s and O6s with Bn Cdr experience at the Army War College. All of the other Bn Cdrs were currently in Bn Cdr positions at FORSCOM locations. In both cases, we were limited to one hour of time. The War College participants did not have any subordinate officers participating in the project, so the time they would have spent providing supervisor ratings was devoted to completing a few additional measures that we could not administer to Bn Cdrs at FORSCOM locations. For this reason, sample sizes for Bn Cdrs vary across predictor and criterion measures.

Table 2.1. *Instruments Administered by Position*

Instruments	PL	CC	Bn S3	Bn XO	Bn Cdr
<i>Predictors</i>					
Figural Reasoning	X		X	X	
Consequences	X	X	X	X	X
Rational Biodata Inventory (RBI)	X	X	X	X	
MCPA-Phrases	X	X	X	X	X
Leader Knowledge Test (LKT)	X	X			
Work Values	X	X			
Virtual Leadership Judgment Simulation ^a	X				
<i>Criteria</i>					
Career History Survey (CHS)	X	X	X	X	
Mental Models Test (MMT)		X			
Technical & Management JKT (PL & CC)	X	X			
Leadership JKT	X	X	X	X	
Provide Performance Ratings for PL		X			
Provide Performance Ratings for CC & BN Staff					X
<i>Maximum Time to Complete Assessments (Hours)</i>	3	3	2	2	1

^a Administered to approximately half of the sample.

Each data collection team comprised two or three members and was led by an experienced data collector. Despite extensive planning, every data collection offered new challenges and the “what if...” section of the Data Collection Manual was updated frequently. Some difficulties were met by workarounds; others were insurmountable. Three of these problems did lead to missing data and are detailed below.

First, we encountered one technical glitch. There was one item on one instrument that, if answered “Yes,” caused the application to terminate and incorrectly indicate completion to the participant. Because there were no obvious clues that the participant had not actually completed the entire predictor-criterion battery, the problem was not identified and corrected until after the first three data collections.

Table 2.2. *Number of Participants by Site*

	PL	CC	Bn XO	Bn S3	Bn Cdr	Not Reported	Total
Army War College	--	--	--	--	27	--	27
Fort Benning	46	27	5	5	6	--	89
Fort Bliss	34	14	3	5	4	--	60
Fort Bragg	37	10	2	2	2	--	53
Fort Carson	48	24	6	5	6	11	100
Fort Drum/Eustis	77	49	9	7	2	4	148
Fort Hood	44	15	1	4	4	1	69
Fort Lewis	57	24	5	6	5	1	98
Fort Riley	82	35	7	10	5	--	139
Fort Stewart	54	21	5	5	6	3	94

Note. Data were collected for Ft. Drum and Ft. Eustis in back-to-back weeks. Data were processed together and thus we were unable to distinguish the two sites.

Second, as noted above, we needed to obtain a supervisor rating for each officer who participated in the project (with the exception of Bn Cdrs). Therefore, our RSRs specified that the chain of command relationship was critical (in order to obtain performance ratings from direct supervisors). However, most locations found it difficult to fully satisfy this requirement. In most cases, units were returning from combat deployments and by the time of Umbrella Week, a number of the unit's officers had already transferred to their next duty assignment. This meant that it was sometimes difficult to locate the higher echelon officer who had directly and recently supervised each participant. When appropriate supervisors of PLs were not immediately available, we gave the PLs instruction cards detailing on-line ratings procedures and asked them to give the instruction card to their former CC. In the case of non-participation by battalion commanders, we made office calls. During the appointment, and depending on how much time the Bn Cdr had available, we provided a laptop or a paper-based rating (PBR) form. We also obtained PBRs if the appropriate supervisor had not been tasked for the full data collection but was willing to come in for a brief period of time, when the appropriate supervisor was the XO or S3, and when Bn Cdrs had to provide ratings for more than six officers.

Finally, while the requirement to provide 12 PLs, per our RSR, was straightforward in maneuver battalions, it was less easily satisfied in artillery and support battalions whose organization does not conform to the maneuver unit structure. Therefore, in some units, we could not achieve our sampling goal for PLs.

Database Development

We combined data across the 11 data collection sites and then split the database by measure to process and score separately. Recall that CCs and Bn Cdrs provided supervisor ratings of their subordinates within the battalion. The supervisor ratings were extracted from the InterForm®-based instruments and combined with supervisor PBR. The rating data were processed and merged with the subordinates' scored data using name and data collection site. OLRDB and ROTC LDAC archival data were merged with the scored survey data and supervisor ratings to create the final database.

Database Screening and Cleaning

We conducted a number of screens to identify careless responding. We did this at the level of individual instruments and across all instruments. For each instrument, each case was marked as “usable” or “unusable” based on screens described below. Then, we created an overall usability screen for each case, based on the number of usable measures.

Data Cleaning Flags

The data set for each instrument was screened for missing data, flat responding, deviance from the mean, and other factors specific to that instrument. An individual’s set of responses on a particular instrument were flagged if: (a) more than 10% of the items had missing values, (b) there was zero or nearly zero variance in the responses (flat responding), or (c) the mean difference between the individual’s response set and the mean response set across all participants was four or more standard deviations above the mean. Finally, we applied screens specific to individual instruments where relevant, for example, response distortion screens for a biodata inventory and screens for illogical response patterns when the instrument involved ranking.

Usability Flag

After determining whether each participant’s data would be “usable” or “unusable” for each instrument, we calculated an overall usability screen. We excluded a participant’s data for all instruments if he/she had more than 50% “unusable” instrument flags. We did make an exception to this rule if there was evidence that the participant responded carefully initially but became fatigued or careless over time. Specifically, if most of the instrument-level “unusable” screens occurred in the second half of the session, then we kept the data for instruments administered in the first half of the session.

Sample Demographics

Of the 877 officers who took the predictor- and criterion-test battery, 836 had usable data on at least some instruments. For any given measure, the percent of usable data ranged from 85% to 98%.

Tables 2.3 through 2.6 provide information about the sample. Table 2.3 shows the breakdown of Army component, pay grade, commissioning source and level of education by position. The majority of the sample was Regular Army and commissioned through ROTC. Twenty percent of the sample was commissioned through OCS, and 19% was commissioned through USMA. Table 2.4 provides branching information by position. Twenty-three different branches were represented, with the majority from combat arms branches, including 28% Infantry, 11% Armor, and 11% Field Artillery. Table 2.5 provides a breakdown of gender, ethnicity, race, and marital status of the sample by position. The majority of the sample was Male (91%) and White (82%), which is not particularly surprising given that more than 50% of the sample came from combat arms branches. Finally, Table 2.6 provides descriptive statistics for time in service and time in grade by position.

Table 2.3. Final Sample by Army Component, Pay Grade, Commission Source, Level of Education, and Position

Variable	Platoon Leader		Company Commander		Battalion XO		Battalion S3		Battalion Commander ^a		Total Sample	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<i>Army Component</i>												
Regular (Active)	461	99.1	217	99.5	38	97.4	48	100.0	66	100.0	830	99.3
Reserve	1	.2			1	2.6					2	.2
<i>Pay Grade</i>												
O-1	120	25.8									120	14.4
O-2	334	71.8	11	5.0			3	6.3			348	41.6
O-3	10	2.2	201	92.2	2	5.1	10	20.8			223	26.7
O-4			4	1.8	36	92.3	35	72.9	1	1.5	76	9.1
O-5									48	72.7	48	5.7
O-6									17	25.8	17	2.0
<i>Commission Source</i>												
US Military Academy	94	20.2	44	20.2	7	17.9	4	8.3	7	10.6	156	18.7
ROTC scholarship	199	42.8	62	28.4	11	28.2	25	52.1	8	12.1	305	36.5
ROTC non-scholarship	44	9.5	44	20.2	12	30.8	10	20.8	11	16.7	121	14.5
Federal OCS – In-service	33	7.1	22	10.1	5	12.8	7	14.6			67	8.0
Federal OCS – Enlistment	74	15.9	38	17.4	2	5.1	1	2.1	1	1.5	116	13.9
National Guard			3	1.4	1	2.6	1	2.1			5	.6
Direct Appointment	2	.4	3	1.4							5	.6
US Air Force Academy or US Naval Academy			1	.5							1	.1
Other	19	4.1	1	.5							20	2.4
<i>Highest Level of Education</i>												
Associate Degree	3	.6									3	.4
College graduate	428	92.0	163	74.8	12	30.8	12	25.0	4	6.1	619	74.0
Post-graduate credit, no degree	13	2.8	14	6.4	4	10.3	3	6.3			34	4.1
Master's degree	18	3.9	38	17.4	21	53.8	32	66.7	23	34.8	132	15.8
Doctorate			1	.5	1	2.6					2	.2
Professional degree	1	.2	1	.5							2	.2
Other	1	.2									1	.1

Note. Total *N* = 836. Platoon Leader *n* = 465, Company Commander *n* = 218, Battalion S3 *n* = 39, Battalion XO *n* = 48, Battalion Cdr *n* = 66. Sample sizes for variables may not add to 100% due to missing data.

^aIncludes 27 current and 39 former Battalion Commanders, *n* = 66. Former Battalion Commanders completed an abbreviated background survey.

Table 2.4. Final Sample by Branch and Position

Variable	Platoon Leader		Company Commander		Battalion XO		Battalion S3		Battalion Commander ^a		Total Sample	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Adjutant General Corps	4	.9	1	.5							5	.6
Air Defense Artillery	15	3.2	8	3.7	1	2.6	1	2.1			25	3.0
Armor	51	11.0	27	12.4	4	10.3	5	10.4	3	4.5	90	10.8
Aviation	13	2.8	4	1.8	1	2.6	1	2.1			19	2.3
Chaplain Corps												
Chemical Corps	4	.9									4	.5
Corps of Engineers	24	5.2	6	2.8	2	5.1	4	8.3			36	4.3
Dental Corps												
Field Artillery	48	10.3	25	11.5	5	12.8	9	18.8	2	3.0	89	10.6
Finance												
Infantry	136	29.2	62	28.4	11	28.2	14	29.2	7	10.6	230	27.5
Judge Advocate General Corps									1	1.5	1	.1
Medical Corps									5	7.6	5	.6
Medical Service Corps	30	6.5	9	4.1	1	2.6	1	2.1			41	4.9
Medical Specialist Corps	1	.2									1	.1
Military Intelligence	18	3.9	4	1.8							22	2.6
Military Police Corps	11	2.4	6	2.8	2	5.1	2	4.2	1	1.5	22	2.6
Nurse Corps	1	.2									1	.1
Ordnance Corps	27	5.8	16	7.3	2	5.1	4	8.3	1	1.5	50	6.0
Quartermaster Corps	26	5.6	21	9.6	6	15.4	2	4.2	1	1.5	56	6.7
Signal Corps	15	3.2	7	3.2							22	2.6
Transportation Corps	37	8.0	18	8.3	3	7.7	4	8.3	1	1.5	63	7.5
Veterinary Corps	1	.2	1	.5							2	.2
Other									5	7.6	5	.6

Note. Total *N* = 836. Platoon Leader *n* = 465, Company Commander *n* = 218, Battalion S3 *n* = 39, Battalion XO *n* = 48, Battalion Cdr *n* = 66. Sample sizes for variables may not add to 100% due to missing data.

^a Includes 27 current and 39 former Battalion Commanders, *n* = 66. Former Battalion Commanders completed an abbreviated background survey.

Table 2.5. Final Sample by Gender, Ethnicity, Race, Marital Status, and Position

Variable	Platoon Leader		Company Commander		Battalion XO		Battalion S3		Battalion Commander ^a		Total Sample	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Gender												
Female	47	10.1	21	9.6	5	12.8	1	2.1	3	4.5	77	9.2
Male	418	89.9	197	90.4	34	87.2	47	97.9	63	95.5	759	90.8
Hispanic												
No, not of Hispanic origin	415	89.2	192	88.1	37	94.9	43	89.6	64	97.0	751	89.8
Yes	39	8.4	21	9.6	2	5.1	4	8.3	2	3.0	68	8.1
Race												
American Indian or Alaska Native	1	.2	2	.9							3	.4
Asian (e.g., Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese)	13	2.8	4	1.8	1	2.6	4	8.3			22	2.6
Black or African American	38	8.2	21	9.6	4	10.3	7	14.6	6	9.1	76	9.1
Native Hawaiian or other Pacific Islander (e.g., Samoan, Guamanian, Chamorro)	3	.6									3	.4
White	375	80.6	180	82.6	33	84.6	35	72.9	58	87.9	681	81.5
Mixed Race	24	5.2	7	3.2	1	2.6	2	4.2	2	3.0	36	4.3
Marital Status												
Single	267	57.4	64	29.4	5	12.8	6	12.5	NA		342	40.9
Married	198	42.6	154	70.6	33	84.6	42	87.5			427	51.1

Note. Total *N* = 836. Platoon Leader *n* = 465, Company Commander *n* = 218, Battalion S3 *n* = 39, Battalion XO *n* = 48, Battalion Commander *n* = 66. Sample sizes for variables may not add to 100% due to missing data.

^a Includes 27 current and 39 former Battalion Commanders, *n* = 66. Former Battalion Commanders completed an abbreviated background survey.

Table 2.6. *Time in Service and Time in Grade for the Final Sample*

Position	Time in Service (Years)		Time in Grade (Months)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Platoon Leader	3.58	3.18	11.67	6.63
Company Commander	8.62	4.01	39.62	15.34
Battalion XO	16.38	4.42	45.68	18.15
Battalion S3	14.31	4.29	37.70	21.50
Battalion Commander ^a	21.41	2.58	--	--
Total Sample	7.52	6.54	22.97	18.25

Note. Total $N = 836$. Platoon leader $n = 465$, Company Commander $n = 218$, Battalion S3 $n = 39$, Battalion XO $n = 48$, Battalion Commander $n = 66$.

^a Includes 39 former Battalion Commanders tested at the War College. They completed an abbreviated background survey with limited items.

CHAPTER 3: MEASUREMENT OF THE CRITERION DOMAIN

Amanda Koch, Bethany H. Bynum, Roy Campbell, William Holden, and Karen Moriarty

We aggregated a wide array of criterion variables from numerous sources to enhance breadth of coverage of the criterion domain. We reduced those variables to a set of meaningful scores using rational and empirical analyses. The final set of criterion scores included one retention criterion (i.e., Career Intentions) and four job performance dimensions: (a) Technical Task Proficiency (TTP); (b) Management, Administration, and Communication Proficiency (MACP); (c) Physical Fitness and Effort (PFE); and (d) Leadership, Supervision, and Personal Discipline (LSPD). Finally, we developed a scheme for weighting the four job performance dimensions differentially across the five positions (PL, CC, Bn XO, Bn S3, and Bn Cdr). This chapter describes criterion measures, the combination of measures into criterion scores, psychometric properties of the criterion scores, and weights developed for use in the validation analyses.

Criterion Measures

Criterion measurement methods were organized into four categories:

- Performance ratings – performance ratings made by supervisors.
- Objectively scored tests – declarative and procedural knowledge tests using a multiple choice (MC) response format.
- Self-report – survey items covering background, experience, performance, attitudes, and career intentions.
- Archival variables – disciplinary actions, awards, promotion rate, and other indicators of achievement.

The development of the criterion measures is described in Paullin et al. (2011). The following paragraphs provide short summaries of each instrument.

Performance Rating Scales

The Performance Rating Scales (PRS) included 13 core items for all positions as well as two items unique to battalion S3 and XO positions, as shown in Table 3.1. The items are based on analyses of officer duty requirements, and are intended to cover a broad range of critical duties performed by officers. Supervisors were instructed to rate one or more subordinate officers and to (a) consider the officer's typical performance over time, and (b) indicate how well the officer performed relative to other officers with similar commissioned experience using the rating scale in Figure 3.1. Scores range from 1 to 5 for each PRS item.

Table 3.1. Performance Rating Scale (PRS) Items

Item	Text
1	Performs branch-specific technical and tactical duties proficiently.
2	Performs core warrior tasks required of all personnel proficiently.
3	Communicates clearly and persuasively in writing.
4	Is effective in oral discourse; listens actively; speaks clearly and persuasively.
5	Demonstrates effort and willingness to keep working under adverse conditions.
6	Demonstrates self-control and personal discipline on the job; provides leader presence and composure.
7	Maintains physical fitness, strength, and weight effectively.
8	Fosters teamwork and enthusiasm for accomplishing objectives; supports and empowers subordinates.
9	Provides structure, direction, training, and instruction to subordinates and informs them of things they should know.
10	Makes sound decisions, innovates solutions to problems, and adapts strategies to changing situations.
11	Plans, coordinates, staffs and monitors unit activities, using resources effectively to accomplish goals.
12	Represents the Army effectively in cross-cultural, multinational, or joint-forces settings.
13	Performs day-to-day administrative tasks, keeping accurate records and reports.
14	Performs battalion S3-specific job duties effectively. (Bn S3 Only)
15	Performs battalion XO-specific job duties effectively. (Bn XO Only)

(1)	(2)	(3)	(4)	(5)	
Well Below Average	<i>Next</i>	Average	<i>Next</i>	Well Above Average	Not Observed
<i>Bottom 20% of Peer Group</i>	<i>20% of Peer Group</i>	<i>Middle 20% of Peer Group</i>	<i>20% of Peer Group</i>	<i>Top 20% of Peer Group</i>	/ Cannot Rate

Figure 3.1. Performance rating scale.

Objectively Scored Job Knowledge Tests (JKTs)

Army-Wide Technical JKT

The Army-Wide (AW) Technical JKT contained seven multiple-choice items aimed at core warrior tasks such as navigation and fieldcraft. Two items were eliminated from analyses because they demonstrated weak item statistics. The examinee's final score was a count of the number of correct responses. A sample item appears in Figure 3.2.

Management JKT

The management JKT was a declarative knowledge test that asked multiple-choice questions relating to troop leading procedures and managing the unit to achieve its goals. It had 23 items; 19 items survived psychometric review. The examinee's final score was a count of the number of correct responses. A sample item appears in Figure 3.2.

Leadership JKT

The Leadership JKT was a measure of leadership judgment in work situations. While it was formatted like a situational judgment test (SJT), it was primarily a measure of procedural knowledge about how to handle leader scenarios. It presented scenarios and asked the respondent to judge the effectiveness of several courses of action for each scenario. There were separate Leadership JKTs for PL, CC, Bn S3, and Bn XO positions. Each JKT had 10 scenarios. The PL version was scored by correlating the examinee's responses with a key based on the responses of 40 transportation and signal officers in the Captain's Career Course (CCC). JKTs for other positions were scored against a peer key, based on the total pool of responses. A few Leadership JKT items were aimed at dimensions other than leadership and were scored on those dimensions as described in the following section on scoring. A sample item appears in Figure 3.2.

AW Technical JKT

Your platoon has been fired on by the enemy. Upon contact your Soldiers have immediately assumed the nearest covered positions. What should your Soldiers do next?

- a. Move out of the engagement area
- b. Conduct an assault
- c. Identify vulnerable flanks
- d. Return fire

Management JKT

Which of the following can be accomplished during the conduct of a mission rehearsal? *Select all that apply.*

- a. Facilitate selection of the best course of action (COA).
- b. Confirm coordination requirements between the company/platoon and adjacent units.
- c. Reinforce performance in critical mission tasks.
- d. Determine the risk assessment.

Leadership JKT

You are a new platoon leader who has just taken charge of a platoon during unit Reset. Many members of the platoon are veterans of at least one combat deployment. Also, you failed to graduate from Ranger School. You are concerned about building credibility with your Soldiers. Please rate the effectiveness of the following actions you could take:

- a. Ask the members of the platoon to share their deployment experiences, and to explain what they learned and how it can help the platoon.
- b. Work hard to get into excellent physical shape so that you excel in PT.
- c. Maintain a strict military posture in all interactions including dress, appearance, grooming, courtesy, reporting, and forms of address, in order to set the example you want followed.
- d. Defer to your Soldiers on matters related to their deployment experiences, acknowledging that they know more than you do in some areas.
- e. Tell your NCOs about all of the studying you have done to increase your competence.
- f. Announce up front that you are in charge and the Soldiers must accept this fact and treat you with appropriate respect.

Figure 3.2. Example Job Knowledge Test (JKT) items.

Mental Models Test (MMT)

The MMT was an experimental measure designed to capture individual differences related to implicit understanding of job requirements. It was administered only to CCs. They were asked to rate the effectiveness of 31 different behaviors when deployed for combat operations using a 10-point rating scale (1 = Extremely Ineffective to 10 = Extremely Effective). Examples of statements include: Devote a lot of time to physical training, and Win an argument. The MMT was scored by computing a C-score, the correlation between the individual's vector of effectiveness ratings and the vector of mean ratings based on data from the full sample of CCs.

Self-Report: Career History Survey (CHS)

The CHS comprised four sections (Background Information, Training and Assignment History, Self-reported Performance, and Career Intentions and Attitudes) with 61 questions. The survey took about 15 minutes to administer to company grade officers (O2-O3). Field grade officers (O4-O5) had four to five extra questions based on their longer time in service (TIS).

Many of the items in the CHS were intended for sample description purposes (e.g., gender, initial branch assignment) but a number of them were linked to indicators of performance or retention (e.g., organizational commitment and career satisfaction). Examples of criterion variables from the CHS appear in Figure 3.3.

Archival Variables

We requested archival records from several databases:

- Officer Longitudinal Research Database (OLRDB), maintained by ARI and updated biennially contains selected demographic and performance data.
- Officer Master File (OMF) and Separation Officer Master File (SOMF) files, maintained by the U.S. Army Human Resources Command (HRC) and provided to ARI on a quarterly basis. ARI uses the OMF and SOMF to update the OLRDB.
- Army Training Requirements and Resources System (ATRRS), which is maintained by the Army G-1, includes information about military coursework.

Data were used to form the following variables:

- *Basic Officer Leader Course – B (BOLC B) recycles and graduation.* The ATRRS database includes basic information on whether an officer recycled (i.e., had to repeat) a portion of the course at least once and whether the officer graduated from the course.
- *Captain's Career Course (CCC) recycles and graduation.* We extracted information from ATRRS about CCC performance for officers who had a rank of captain or higher.

Training and Assignment History

How often have you served as a formal briefing officer to higher command echelons?

Self-Reported Performance

- Were you ranked in the top 20% of your class in BOLC B/BOLC II?
- *For Company Commanders, Battalion XO's, Battalion S3's, and Battalion Commanders only:* Were you ranked in the top 20% of your class in CCC (Advanced Course)?
- Have you ever been formally counseled about your behavior or discipline?
- What was your latest Army Physical Fitness Test (APFT) test score?
- Has the senior officer or rater ever recommended you for command on the OER?
- How many of your OER ratings have been above center of mass?
- As an officer, how many times have you been selected for below-the-zone promotion?

Career Intentions and Attitudes

Which of the following best describes your current active duty career intentions?

- I plan to stay in the Army beyond 20 years.
- I plan to stay in the Army until retirement (e.g. 20 years or when eligible to retire).
- I plan to stay in the Army beyond my obligation, but I am undecided about staying until retirement.
- I am undecided whether I will stay in the Army upon completion of my obligation.
- I will probably leave the Army upon completion of my obligation.
- I will definitely leave the Army before I reach my retirement point.

Figure 3.3. Example Career History Survey (CHS) items.

- *Total Awards.* OLRDB and OMF data provide information about awards. Officers can be awarded medals for many reasons, such as deploying (e.g., Afghanistan Campaign Medal), serving on active duty (e.g., National Defense Service Medal), and meritorious service (e.g., Army Commendation Medal). We constructed the awards variable in two ways. First, we created a variable that was the simple sum of the meritorious service/valor. Second, recognizing that some medals are more difficult to achieve than others, we created a second variable that differentially weighted the meritorious service/valor awards. We derived the weights from the Promotion Points Worksheet for Noncommissioned Officers (Department of Army, 2010). Awards variables were standardized within branch to remove any branch differences in opportunity.
- *Promotion Variables.* Using OLRDB and OMF data, we constructed promotion rate variables for Promotion to First Lieutenant and Promotion to Captain by counting the number of days between promotions and standardizing within branch to remove branch differences. After standardizing within branch we also standardized within commissioning years to account for the force increase that took place between 2003-2010, because time to promotion was likely shorter during that time period. Promotion to First Lieutenant occurs for virtually all officers and is mostly based on time in service, so we did not expect a great deal of variability across PLs on this

criterion measure. In contrast, promotion to captain has become more competitive in recent years, so we anticipated that there could be variability across CCs, and that time to promotion could be a reflection of their performance.

Development of the Career Intentions Score

The Career Intentions score was derived by computing a standardized average of three CHS self-report items about career intentions.

Development of Performance Dimension Scores

Given the large number of criterion variables, we sought a way to meaningfully combine criterion items and scales into broader criterion dimensions. A performance model developed in prior ARI research (Campbell, 2012; Campbell & Knapp, 2001; Campbell, McCloy, Oppler, & Sager, 1993) identified eight performance components present in all jobs (branch-specific technical proficiency; Army-wide technical proficiency; written and oral communication task proficiency; demonstrating effort; maintaining personal discipline; maintaining physical fitness, strength, and weight; leadership and supervision; and management and administration). In our present data set, we used a content-oriented approach, informed by empirical analyses, to sort the criterion items and scales into these criterion components. Subject Matter Experts (SMEs) from ARI and HumRRO made ratings to indicate how well each criterion item or scale fit into each component. Disagreements about ratings were resolved through discussion.

Based on empirical analyses, we dropped a few criterion items in one or more position-specific samples¹ due to (a) low item-total correlations for the measure, (b) negative correlations between an item and other items linked to the same component, (c) low base rates or missing data, or (d) inappropriateness of an item for a particular sample. Table 3.2 presents the preliminary performance components that were carried forward for the next set of analyses. The components mirror the original performance model with one exception. Two components, branch-specific task proficiency and Army-wide technical task proficiency, were combined to create a single technical task proficiency component because several criterion items/scales were linked (based on rater judgments and empirically) to both components.

Next, we conducted a series of Confirmatory Factor Analyses (CFAs) using the largest sample (PL) to (a) compare the 7-component model to other potential substantive models and (b) examine common method variance.² Fit indices were compared for models with 1, 3, 4, 5, 6, and 7 performance factors as well as models that included these performance factors plus three method factors (performance ratings, objectively scored tests, and self-report factors). Common method variance, represented by the three method factors, accounted for a good deal of variance, and models accounting for method variance fit the PL data much better than those that did not.

¹ While we sought to develop a common performance framework across all positions (PL, CC, Bn XO, Bn S3, and Bn Cdr), the specific criterion measures administered to officers in each position was somewhat different. Therefore, we had to map the criterion items administered to each position-specific sample to the performance framework.

² Because career intentions is an attitudinal rather than a performance-related outcome, it was not included in the CFAs, which were intended to model performance.

Table 3.2. Preliminary Performance Components

Dimension	Definition
1. Technical Task Proficiency	The degree to which the officer performs the core substantive or technical tasks and duties, including those that are central to his or her branch as well as those that are Army-wide.
2. Written and Oral Communication Task Proficiency	The proficiency with which the officer can write or speak, independent of the correctness of the subject matter.
3. Demonstrating Effort	The consistency of an officer's effort day by day, the frequency with which he will expend extra time when required, and the willingness to keep working under adverse conditions. It is a reflection of the degree to which individuals commit themselves to all job tasks, work at a high level of intensity, and keep working when it is cold, wet, or late.
4. Maintaining Personal Discipline	The degree to which the officer avoids negative behaviors, such as alcohol and substance abuse at work, law or rules infractions, and excessive absenteeism.
5. Maintaining Physical Fitness, Strength, and Weight	The extent to which the officer meets or exceeds the Army's standards for fitness.
6. Leadership and Supervision	An officer's use of direct interpersonal interaction to influence the behavior of other people such that their performance is enhanced, both individually and collectively. Includes: encouraging, supporting, empowering, and training/coaching subordinates, influencing own supervisors, and serving as a role model.
7. Management and Administration	An officer's use of (i.e., management of) the unit's/Army's resources to achieve its goals. Includes: articulating goals for the unit, organizing people and resources, monitoring progress, controlling expenditures, and representing the unit in dealings with other units, organizations, or the public. Administration includes performing day-to-day administrative tasks, keeping accurate records, documenting actions, analyzing routine information, and making information available in a timely manner.

The best fitting model included seven performance factors corresponding to our initial mapping and three method factors (see Table 3.3). We determined that we could not use this model for subsequent analyses because several of the performance factors were highly correlated (the same was true for models with 5 or 6 factors), and thus we could not compute factor scores for a 7-factor model. Producing factor scores was an important consideration for model selection because (a) method variance made a substantial contribution to the models and (b) we wanted to use factor scores in subsequent analyses to account for the effects of method variance. The models with 1, 3, and 4 performance factors did produce factor scores.

We retained the model with four performance factors (Technical Task Proficiency; Management, Administration, and Communication Proficiency; Physical Fitness and Effort; Leadership, Supervision, and Personal Discipline) and three method factors for all subsequent analyses because it was the best fitting model that would produce dimension scores.

Table 3.3. Confirmatory Factor Analyses (CFA) Results

Number of Performance Factors	Performance Components ^a	Method Factors ^b	Fit Indices				Change in RMSEA	
			RMSEA	CFI	TLI	WRMR	Reduction	% Improvement
1	1+2+3+4+5+6+7	No	.075	.666	.629	1.373	--	--
3	1+2+7, 3+4+5, 6	No	.073	.686	.645	1.342	.002	2.7
4	1, 2+7, 4+6, 3+5	No	.072	.706	.662	1.306	.003	4.0
7	1, 2, 3, 4, 5, 6, 7	No	.068	.756	.697	1.217	.007	9.3
5	1, 2+7, 3, 4+6, 5	No	.068	.743	.700	1.238	.007	9.3
6	1, 2+7, 3, 4, 5, 6	No	.068	.746	.695	1.233	.007	9.3
6	1, 2, 3, 4+6, 5, 7	No	.067	.752	.702	1.223	.008	10.7
0		Yes	.051	.847	.827	1.198	.024	32.0
3	1+2+7, 3+4+5, 6	Yes	.028	.960	.949	.725	.047	62.7
1	1+2+3+4+5+6+7	Yes	.028	.957	.947	.739	.047	62.7
4	1, 2+7, 4+6, 3+5	Yes	.026	.966	.956	.699	.049	65.3
5	1, 2+7, 3, 4+6, 5	Yes	.023	.974	.966	.662	.052	69.3
6	1, 2+7, 3, 4, 5, 6	Yes	.023	.975	.966	.655	.052	69.3
7	1, 2, 3, 4, 5, 6, 7	Yes	.017	.986	.980	.611	.058	77.3
6 ^c	1, 2, 3, 4+6, 5, 7	Yes	--	--	--	--	--	--

Note. CFA analyses included only the platoon leader sample. RMSEA = Root Mean Square Error of Approximation; acceptable values are less than .08. CFI = Comparative Fit Index; acceptable values are greater than or equal to .95; TLI = Tucker-Lewis Index; acceptable values are greater than or equal to .95. WRMR = Weighted Root Mean Residual; acceptable values are less than .90 (see Hu & Bentler, 1999, for rules of thumb for fit indices).

^a See Table 3.2 for full titles and definitions of the performance components. A "+" indicates that performance dimensions were combined into one factor.

^b Indicates whether three method factors (performance ratings, objectively scored tests, and self-report) were included in the model.

^c Model did not converge.

Each officer ultimately could have a criterion score on four performance factors and three method factors. As noted above, the items and scales comprising the criterion dimensions varied by position. They were highly similar for the PL and CCs, so a model including PLs and CCs was used to derive dimension scores for these two groups. The items and scales were also highly similar for the Bn S3 and Bn XO positions, so a model including Bn S3s and Bn XOs was used to derive factor scores for these two groups. Table 3.4 shows the items and scales that loaded on each of the final factors for each of these two groups. This table also shows the items and scales that were mapped onto the Career Intentions outcome variable, even though these items were not part of the CFAs. The archival variables (such as promotion rate) were not organized into the five main criterion scores, so they are not listed in this table.

Psychometric Properties of Criterion Scores

Tables 3.5, 3.6, and 3.7 provide descriptive statistics, intercorrelations and subgroup differences for the four performance dimension scores and the Career Intentions score. The Career Intentions composite tended to have small (but often significant) correlations with the performance dimensions.

Correlations among the four final performance dimensions were generally large. One reason could be that each dimension includes supervisor ratings and these are often susceptible to rater errors, such as a tendency to give ratees all high (or all low) ratings without much regard to ratee's performance on individual dimensions. Removing method variance should have minimized between-dimension correlations due to rater error. Indeed, this is one reason we removed method variance from the performance factor scores. Still, even after removing method variance, some of the correlations are large, and there is no obvious explanation for why this occurred. For example, the Technical Proficiency factor score is highly correlated with all of the other dimension scores, including the Physical Fitness and Effort score. Also, the significant correlation between Physical Fitness and Effort and the Leadership, Supervision, and Personal Discipline Factor is puzzling. One possible explanation is that Physical Fitness and Effort may be easier to observe and understand than some of the other dimensions. Raters may have relied heavily on their subordinate's easy-to-observe physical prowess when making their ratings on other dimensions. This is an area for future research.

Subgroup differences on the performance dimensions were typically small to medium (see Table 3.7). Black and Hispanic officers had higher scores on the Career Intentions composite than did White, non-Hispanic officers. White, non-Hispanic, and male officers had higher scores than Black and Hispanic officers on the performance factors. Appendix A contains tables that provide descriptive statistics, intercorrelations, and subgroup differences for all of the criterion variables, including archival variables.

Table 3.4. *Final Criterion Dimensions and Measures*

Criterion Dimension	Measures Included for Each Position			
	Platoon Leader & Company Commander	Factor Loadings	Battalion S3 & Battalion XO	Factor Loadings
1. Technical Task Proficiency	• Performs branch-specific technical and tactical duties proficiently	.494	• Performs branch-specific technical and tactical duties proficiently	.526
	• Performs core warrior tasks required of all personnel proficiently	.549	• Performs core warrior tasks required of all personnel proficiently	.557
	• Ranked in top 20% of BOLC B/BOLC II class	.545	• Ranked in top 20% of BOLC B/BOLC II class	.578
	• Army-Wide Declarative Knowledge test: Army-Wide Technical Knowledge Scale	.148	• Ranked in top 20% of CCC class	.686
	• Ranked in top 20% of CCC class	.717		
2. Management, Administration, and Communication Proficiency	• Communicates clearly and persuasively in writing	.371	• Communicates clearly and persuasively in writing	.375
	• Is effective in oral discourse; listens actively; speaks clearly and persuasively	.415	• Is effective in oral discourse; listens actively; speaks clearly and persuasively	.374
	• Makes sound decisions, innovates solutions to problems, and adapts strategies to changing situations	.440	• Makes sound decisions, innovates solutions to problems, and adapts strategies to changing situations	.440
	• Plans, coordinates, staffs, and monitors unit activities, using resources effectively to accomplish goals	.440	• Plans, coordinates, staffs, and monitors unit activities, using resources effectively to accomplish goals	.641
	• Performs day-to-day administrative tasks, keeping accurate records and reports	.237	• Performs day-to-day administrative tasks, keeping accurate records and reports	.415
	• Frequency of serving as a formal briefing officer to higher command echelons	.120	• Frequency of serving as a formal briefing officer to higher command echelons	.280
	• Army-Wide Leadership Test: Management and Administration Knowledge Scale	.119	• Army-Wide Leadership Test: Management and Administration Knowledge Scale	.066
	• Army-Wide Declarative Knowledge Test: Management and Administration Knowledge Scale	.412		
3. Physical Fitness and Effort	• Demonstrates effort and willingness to keep working under adverse conditions	.365	• Demonstrates effort and willingness to keep working under adverse conditions	.376
	• Maintains physical fitness, strength, and weight effectively	.681	• Maintains physical fitness, strength, and weight effectively	.644
	• Most recent Army Physical Fitness Test (APFT) score	.662	• Most recent Army Physical Fitness Test (APFT) score	.629

(continued)

Table 3.4. Final Criterion Dimensions and Measures (Continued)

Criterion Dimension	Measures Included for Each Position			
	Platoon Leader & Company Commander	Factor Loadings	Battalion S3 & Battalion XO	Factor Loadings
4. Leadership, Supervision, and Personal Discipline	• Demonstrates self-control and personal discipline on the job; provides leader presence and composure	.451	• Demonstrates self-control and personal discipline on the job; provides leader presence and composure	.400
	• Fosters teamwork and enthusiasm for accomplishing objectives; supports and empowers subordinates	.370	• Fosters teamwork and enthusiasm for accomplishing objectives; supports and empowers subordinates	.327
	• Provides structure, direction, training, and instruction to subordinates and informs them of things they should know	.337	• Provides structure, direction, training, and instruction to subordinates and informs them of things they should know	.387
	• Formally counseled about behavior or discipline	.139	• Formally counseled about behavior or discipline	.073
	• Recommended for command on the OER	.449	• Recommended for command on the OER	.437
	• Recommended for XO on the OER	.337	• Recommended for XO on the OER	.346
	• Recommended for S3 on the OER	.305	• Recommended for S3 on the OER	.345
	• Army-Wide Leadership Test: Leadership and Supervision Knowledge Scale	.077	• Rated as having general officer potential by the senior officer on the OER	.578
			• Number of OER ratings above center of mass (0-6)	.455
			• Number of times selected for a below-the-zone promotion (0-3)	.192
			• Army-Wide Leadership Test: Leadership and Supervision Knowledge Scale	.073
			• Army-Wide Leadership Test: Self-Discipline	-.086
5. Career Intentions	• Confidence in staying in the Army at least 20 years		• Confidence in staying in the Army at least 20 years	
	• Confidence in staying in the Army at least 25 years		• Confidence in staying in the Army at least 25 years	
	• Confidence in staying in the Army at least 30 years		• Confidence in staying in the Army at least 30 years	
	• Current active duty career intentions		• Current active duty career intentions	
	• Submission of a request release from active duty (REFRAD)		• Submission of a request release from active duty (REFRAD)	

Note. Career Intentions was not included in the factor model and therefore does not have factor loadings. Bolded values indicated significant factor loadings at $p < .05$.

Table 3.5. Criterion Score Descriptive Statistics

Criterion Dimension	Platoon Leader		Company Commander		Battalion S3		Battalion XO	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1 Technical Task Proficiency	-.12	.71	.23	.72	.29	.77	.37	.63
2 Management, Administration, and Communication Proficiency	-.13	.64	.25	.62	.39	.80	.55	.58
3 Physical Fitness and Effort	-.06	.76	.11	.80	.17	.70	.08	.74
4 Leadership, Supervision, and Personal Discipline	-.14	.67	.27	.67	.38	.83	.49	.65
5 Career Intentions	.14	.72	.48	.59	.85	.40	.95	.37

Note. PL $n = 466$. CC $n = 217$. Bn S3 $n = 48$. Bn XO $n = 39$. Performance factors were standardized with a mean of zero and a standard deviation of 1.00 on the full sample. Negative means indicate lower scores on the criterion variable. As shown, many of the PL means are negative, probably because they are less experienced than the higher ranking officers.

Table 3.6. Criterion Score Intercorrelations

Criterion Dimension	1	2	3	4	5
1 Technical Task Proficiency					
2 Management, Administration, and Communication Proficiency	.89				
3 Physical Fitness and Effort	.80	.48			
4 Leadership, Supervision, and Personal Discipline	.98	.91	.76		
5 Career Intentions	.08	.11	.01	.11	

Note. $n = 770$. Bold = $p < .01$.

Table 3.7. Criterion Dimension Subgroup Differences

Criterion Dimension	Male-Female <i>d</i>	WNH-Black <i>d</i>	WNH-Hispanic <i>d</i>
1 Technical Task Proficiency	.30	.51	.20
2 Management, Administration, and Communication Proficiency	.20	.58	.29
3 Physical Fitness and Effort	.26	.23	.07
4 Leadership, Supervision, and Personal Discipline	.31	.47	.22
5 Career Intentions	.15	-.61	-.24

Note. Positive d -values indicate higher scores for the first group listed, and negative d -values indicate higher scores for the second group listed. Male $n = 696$; Female $n = 74$. White Not Hispanic (WNH) $n = 569$; Black $n = 70$. Hispanic $n = 47$.

Development of the Overall Performance Composite

To create an overall performance composite, we developed weights for each of the four performance dimensions for each position. Twenty-two officers (O5s and O6s) at the National Defense University (NDU) were asked to make four judgments:

- (a) The importance, on a 7-point scale ranging from *Somewhat Important* to *Extremely Important*, of each of the original seven performance components (see Table 3.2) for effective performance in each of the five positions (in a deployment context);³
- (b) Whether each performance component becomes more important, less important, or stays the same in importance as rank increases;
- (c) Whether each performance component becomes more important, less important, or stays the same in importance in a garrison context compared to a deployment context; and
- (d) Which three performance components are most important for each of the five positions.

Because a 7-component performance model was used in these expert judgment exercises, we averaged some component ratings to create final ratings that map to the four final performance dimensions (e.g., Branch-specific and Army-wide technical proficiency).

We used the ratings to develop a weighting scheme that would capture the (a) key performance dimensions for each position and (b) important differences between positions. Because all mean importance ratings for the first exercise were high, all performance dimensions were identified as important for all positions (see Table 3.8). Therefore, each performance dimension received a weight in the overall performance composite for each position. A weight of “3” was assigned to performance dimensions that had the highest or second highest mean rating for a position. A weight of “2” was assigned to performance dimensions that did not have the highest or second highest mean rating for a position but did appear in the top three dimensions for a position in the expert judgment exercise (select the three most important performance dimensions). A weight of “1” was assigned to the performance dimension that did not meet the other decision rules.

Table 3.8. Performance Dimension Mean Ratings (*M*) and Weights (*W*)

Criterion Dimension	PL		CC		Bn XO		Bn S3		Bn Cdr	
	<i>M</i>	<i>W</i>	<i>M</i>	<i>W</i>	<i>M</i>	<i>W</i>	<i>M</i>	<i>W</i>	<i>M</i>	<i>W</i>
Technical Task Proficiency	5.91	3	6.05	2	5.23	1	5.50	1	5.18	1
Management, Administration, and Communication Proficiency	4.45	1	5.39	1	6.30	3	6.07	3	6.66	3
Physical Fitness and Effort	6.25	3	6.30	3	5.77	1	5.86	1	6.05	1
Leadership, Supervision, and Personal Discipline	5.77	2	6.25	3	6.18	3	6.09	3	6.82	3

Note. Ratings were made on a 7-point scale where 1 = Somewhat Important, 3 = Important, 5 = Very Important, and 7 = Extremely Important.

The weights in Table 3.8 were used to create an overall performance composite for each officer in each position. The weights assigned to performance dimensions were the same for Bn XO's, Bn S3s, and Bn Cdrs, so the overall performance composite was derived in the same way for

³ Over the course of several projects, we have found that officers typically ask us whether importance ratings should be made for a deployment setting or a garrison setting, and we need to provide context for them to use in making judgements. We instructed SMEs to use a deployment setting as a frame of reference.

these three positions. However, the weights do reflect some differences between more junior and senior positions. Namely, the Technical Task Proficiency and Physical Fitness and Effort dimensions are more important for platoon leaders and company commanders, whereas the Management, Administration, and Communication Proficiency dimension is more important for the three battalion-level positions.

CHAPTER 4: PREDICTOR MEASURES AND PSYCHOMETRICS

Cheryl Paullin, Michael Cullen, and Justin Arneson

Predictor measures were selected to tap cognitive and non-cognitive constructs thought to be important for effective performance and retention in officer positions. This chapter describes predictor measures and their psychometric properties based on data from the concurrent validation project.

Predictor Measures

The research team worked with the technical advisory committee to identify (a) existing measures from operational programs and experimental research efforts that could be used or reworked to measure skills, abilities, and other characteristics (SAOs) identified in the job analysis (Paullin et al., 2011) and (b) other high-potential measures that would add value to the predictor set for a validation project. Table 4.1 maps the selected measures against predictor constructs.

Table 4.1. *Predictor Measurement Plan*

Measures	Predictor Constructs						
	General Cognitive Ability	Temperament	Physical/ Psychomotor	Applied Cognitive Skills	Social/ Leadership Orientation	Communication Skills	Interests/ Values
Figural Reasoning	X						
Essay Test						X	
Objective-Format Consequences Test	X			X			
Rational Biodata Inventory (RBI)		X			X	X	
Modified Career Path Appreciation (MCPA)- Phrases				X			
Leader Knowledge Test (LKT)					X		
Work Values							X
Virtual Leadership Judgment Simulation (VLJS)				X	X		
ASVAB/ SAT/ACT scores	X						
Undergraduate GPA	X						
ROTC/OCS/USMA OML			X	X	X		X
ROTC LDAC Performance			X	X	X		X

Figural Reasoning Test

The Figural Reasoning test is a 20-item measure of non-verbal reasoning. It was originally developed as a part of the U.S. Army's Project A (Campbell & Knapp, 2001) and was administered as part of the joint-service Enhance Computer Administered Testing (ECAT) project (Alderton, Wolfe & Larson, 1997). When presented via computer, each item comprised a series of four figures shown on-screen at once. The examinee must induce the rule that explains the serial changes in the figures and select the figure, out of five choices, that would be the next one in the series according to that rule. The score on Figural Reasoning was a count of the number of correct responses.

Rational Biodata Inventory (RBI)

The RBI uses rational biodata items to measure temperament attributes. The test was developed by ARI and used in prior ROTC and OCS (Allen et al., 2012; Putka, 2009, Russell & Tremble, 2011) projects. It used a 5-point Likert-type response format (e.g., strongly agree to strongly disagree or very often to never). Scales chosen for inclusion in the predictor measure for the current project are described in Table 4.2.

When examining RBI data from ROTC and OCS projects, we found that one item about doing pull-ups resulted in a large subgroup difference between males and females on the Fitness Motivation scale. Upon further examination, items that asked respondents to compare themselves to others of the same sex and age yielded smaller subgroup differences than items that did not provide a referent group. Items that addressed cardiovascular fitness appeared to have lower male-female differences than items about strength. In attempt to minimize male-female subgroup differences while maintaining validity, we wrote new items, where appropriate, using the referent group designation. We also added new items targeting Hogan's (1991) three factors of physical ability—strength, cardiovascular endurance, and flexibility and balance. Including items for all three factors allowed for a more comprehensive assessment of fitness motivation while potentially mitigating the traditionally high gender differences in strength with flexibility and balance where there is often a female advantage. We constructed three fitness motivation scores: (a) the original score that included the pull-up item and has been used in ROTC and OCS projects, (b) an alternative score that excluded the pull-up item and included the new strength, cardiovascular endurance, and flexibility and balance items, and (c) an alternative score that excluded the pull-up item and included the new strength and cardiovascular endurance items. The scores were compared to determine if the alternative scores reduced subgroup differences and maintained validity.

Modified Career Path Appreciation (MCPA)-Phrases

The MCPA-Phrases assessment was a product of ARI's research on complex cognitive skills needed for successful strategic leadership performance (Jacques & Stamp, 1990; Lewis, 1993). It was originally a small part of a larger interview. High scores reflect a greater tolerance for working in ambiguous, complex and uncertain environments which, according to Stratified Systems Theory, are more closely associated with senior leadership positions.

Table 4.2. *Rational Biodata Inventory (RBI) Scale Definitions*

Scale	<i>k</i>	Definition
Army Affective Commitment	9	The degree of personal identification with, and intrinsic interest in being, a U.S. Army Soldier. The extent to which a candidate feels emotionally attached to the Army.
Achievement Orientation	6	The willingness to give one's best effort and to work hard towards achieving difficult objectives.
Hostility to Authority	5	Belief that superiors abuse their power and take advantage of their employees. This scale was taken from another ARI instrument, the ARC.
Career Expectations	7	Desire and intent to remain in the Army.
Fitness Motivation – All Items	14	Degree of enjoyment from participating in physical exercise. Willingness to put in the time and effort to maintain good physical conditioning.
Fitness Motivation – Original (Old)	8	Fitness motivation scale used in prior ROTC work. Emphasizes strength.
Fitness Motivation – New Composite 1 (New 1)	12	Fitness items targeted toward general fitness, strength, endurance, and flexibility/balance.
Fitness Motivation – New Composite 2 (New 2)	9	Fitness items targeted toward general fitness, strength, and endurance.
Goal Orientation	6	Has a high expectation of future achievements relative to other officers.
Tolerance for Injury	6	Tolerance for situations where risk of injury is possible. Attraction to activities involving risk of injury or operating in a dangerous setting.
Peer Leadership	6	Seeks positions of authority and influence. Comfortable with being in charge of a group. Willing to make tough decisions and accept responsibility for the group's performance.
Self-Efficacy	6	Feeling that one has successfully overcome work obstacles in the past and that one will continue to do so in the future.
Tolerance for Ambiguity	7	Ability to tolerate work situations where the right goal or the correct path to the goal is vague and ill-defined.
Written Communication	7	The ability to clearly communicate one's ideas in writing to others.

Note. All items use a 5-point Likert rating scale.

This version of the MCPA-Phrases (Paullin et al, 2011), contained 9 sets of 6 behaviors. Examinees were asked to rank order the behaviors in each set in terms of how well the behaviors reflected the examinee's "typical way of behaving." MCPA-Phrases yielded two scores: (a) Career Path Approach (CPA) potential score and (b) Profile Similarity Index (PSI). CPA potential was the sum of scores for behaviors ranked "1" (most like "typical way of behaving") by the examinee. PSI was the average Spearman rank-order correlation for each item set, where each correlation reflected the similarity between the respondent's response to a set and the score profile for those items based on the CPA scoring key.

Leader Knowledge Test (LKT)

The LKT is a measure of implicit leadership theories and quantifies the extent to which the examinee's beliefs and assumptions about officer requirements match those of other officers (Legree, Kilcullen, Putka, & Wasko, 2014). ARI developed and tested the LKT in prior ROTC

(Legree, Kilcullen, Psotka, Putka & Gintner, 2010) and OCS projects (Allen et al., 2012; Russell & Tremble, 2011). In this project, the LKT was only administered to Platoon Leaders and Company Commanders. Examinees were asked to rate the importance of each characteristic and skill for being a successful Army Company Grade leader (i.e., a platoon leader or a company commander) using a 10-point importance rating scale. We computed six LKT scores—C-Scores, elevation, and scatter scores for both the skills and characteristics, where the C-Score is the correlation between the examinee's responses and the vector of mean responses for all examinees and elevation and scatter are the mean and variance of examinee responses.

Work Values

Over the course of prior ROTC (Putka, 2009) and OCS (Allen et al., 2012; Russell & Tremble, 2011) projects, ARI and HumRRO developed a tool for evaluating work values and fit between the person's values and those provided by the job. In this project, examinees were asked to (a) rank order 11 job characteristics in terms of how important they would be to have on an ideal job and (b) select the characteristics that would need to be present in order for a job to be ideal. We created a score for each examinee on each of the 11 job characteristics based on the combination of their two judgments. We also calculated a Profile Similarity Index (PSI) score for each examinee. It was the Spearman rank order correlation between the examinee's rank ordered values and a profile of values as ranked by a sample of Army captains in a separate project.

Objective-Format Consequences Test

The Consequences Test (Christensen, Merrifield, & Guilford, 1953) is a measure of divergent thinking that was recently shown to predict long-term continuance and performance among Army officers (Zaccaro et al., 2012). It has an open-ended response format. Given its promise, we investigated alternative objective response format versions because we knew it would be administratively cumbersome for the Army to administer an open-ended test operationally. The final version contained five scenarios, each with 17 possible consequences, to total 85 items. Examinees were asked to rate the originality of each consequence using the following 9-point rating scale:

- 1-3: not very original, most people would think of this consequence
- 4-6: somewhat original, some people would think of this consequence
- 7-9: highly original, few people would think of this consequence

Two scores were computed: (a) the deviation between the examinees's originality rating and mean ratings from the combined sample of PLs and CCs and (b) the C-Score, the correlation between the examinee's originality rating and the vector of mean originality ratings made by the combined sample of PLs and CCs.

Virtual Leadership Judgment Simulation (VLJS)

The VLJS was a measure of leadership judgment that was developed as a part of this project. The VLJS asked the examinee to assume the role of Team Leader of an ROTC company. The examinee received background information about the team and information about the task at

hand. Animated scenarios provided additional information to help examinees make decisions about the task. After each scenario, the examinee answered 40 items that were developed to assess leader judgment and decision making. Thirty-one of the items were scored on a Likert-type scale and the other 9 items used absolute rating scales. Six scores were computed for the VLJS:

- C-score correlating examinee's responses with the vector of median responses from all examinees
- C-score correlating examinee's responses with the vector of mean responses from all examinees
- C-score correlating examinee's responses with the vector of median responses from a sample of Professors of Military Science
- C-score correlating examinee's responses with the vector of mean responses from a sample of Professors of Military Science
- Deviation score between examinee responses and responses from all examinees
- Deviation score between examinee responses and responses from a sample of Professors of Military Science.

College GPA Category

While we planned to obtain GPA from archival records, we knew we would not be able to obtain it for much of the sample. Therefore, in the CHS, participants were asked: "What was your undergraduate GPA?" The response scale had four categorical items: (a) 2.0 – 2.5, (b) 2.6 – 3.0, (c) 3.1 – 3.5, and (d) 3.5 or above, which were scored 1 through 4, respectively.

Archival Predictor Variables

For officers commissioned through ROTC, the U.S. Army Cadet Command provided archival records of (a) rank on the Order of Merit List (OML) at the end of ROTC, (b) GPA in college, and (c) performance in the Leader Development and Assessment Course (LDAC). For a smaller subset of ROTC-commissioned officers who were awarded 4-year ROTC scholarship, we obtained SAT/ACT scores and Whole Person Scores that Cadet Command uses to award scholarships.⁴

Psychometric Properties of Predictor Scores

As shown in Table 4.3, almost all the reliability estimates for the predictor scores were adequate and several were quite high. The Objective-Format Consequences test score reliability estimates were over .90, which is not surprising because the test has 85 items. The reliabilities for the MCPA-Phrases measure were disappointingly low. Quality checks for all of the keys and scoring steps did not reveal any reason for the poor reliability; however, items in this version of

⁴ The Whole Person Score is a composite of score on cognitive and non-cognitive assessments.

the MCPA-Phrases were presented in an order different from the order used in the original research, so possibly this impacted reliability.

Table 4.4 provides effect sizes for subgroup differences. The most striking effect size is for the Objective-Format Consequence Test that yields a large subgroup difference. Whites score approximately one standard deviation higher than Blacks. Also, for the Fitness Motivation – Old composite, females score about one standard deviation lower than males. The new Fitness Motivation scales yield much smaller male-female subgroup differences. Too few Hispanics participated in the project to permit comparison of scores from Hispanics to those from other subgroups.

Appendix B contains two additional predictor tables. Table B.1 contains predictor score means by position, and Table B.2 presents the full predictor intercorrelation matrix.

Table 4.3. *Predictor Score Descriptive Statistics*

Predictor	<i>n</i>	<i>k</i>	<i>MN</i>	<i>SD</i>	Alpha
Figural Reasoning	448	20	14.679	3.493	.753
Objective-Format Consequences					
Deviation Score	634	85	7.038	.633	.962
C-score	644	85	.440	.338	.918
Rational Biodata Inventory (RBI)					
Army Affective Commitment	659	9	3.723	.652	.810
Achievement Orientation	659	6	4.066	.492	.680
ARC Hostility	659	5	2.823	.719	.770
Career Expectations	659	7	3.096	.986	.910
Fitness Motivation – All Items	659	14	3.798	.516	.840
Fitness Motivation – Old	682	8	3.923	.576	.794
Fitness Motivation – New 1	682	12	3.792	.508	.812
Fitness Motivation – New 2	682	9	3.908	.542	.799
Goal Orientation	659	6	3.459	.644	.740
Tolerance to Injury	659	6	3.900	.592	.690
Peer Leadership	659	6	3.683	.602	.810
Self-Efficacy	659	6	4.289	.454	.780
Tolerance for Ambiguity	659	7	3.210	.519	.760
Writing	659	7	3.331	.590	.750
MCPA-Phrases					
Raw Total	631	9	3.579	.570	.143
Adj. Raw Total	631	9	12.538	2.347	.143
Profile Similarity Index	631	9	.029	.141	.266
Leader Knowledge Test (LKT)					
Characteristics C-score	635	30	.773	.153	.751
Skills C-score	632	30	.696	.182	.646
Characteristics Elevation	636	30	7.058	.838	.879
Skills Elevation	636	30	7.367	1.024	.930
Characteristics Scatter	636	30	6.102	2.632	.796
Skills Scatter	636	30	4.162	2.549	.851

(continued)

Table 4.3. Predictor Score Descriptive Statistics (Continued)

Predictor	<i>n</i>	<i>k</i>	<i>MN</i>	<i>SD</i>	Alpha
Work Values					
Do work that challenges me	655	1	-.015	.936	--
Work in a comfortable, relaxed environment	655	1	-.130	1.006	--
Do work that keeps me close to home	655	1	-.135	.952	--
Provide guidance and direction to others	655	1	.047	1.071	--
Receive a good salary and benefits	655	1	.017	1.018	--
Receive recognition or praise for what I do	655	1	-.067	.999	--
Come up with my own way to do tasks	655	1	-.065	.929	--
Contribute to society and the well-being of others	655	1	.593	1.152	--
Have well-defined rules for accomplishing tasks	655	1	-.226	1.007	--
Work as part of a team	655	1	-.018	.907	--
Work on a variety of types of problems	655	1	.181	.881	--
Profile Similarity Index	655	12	.084	.328	^a
Virtual Leadership Judgment Simulation					
C-score consensus (using medians)	248	40	64.104	5.289	.610
C-score consensus (using means)	248	40	64.537	4.955	.570
C-score SME (using medians)	248	40	62.860	5.387	.520
C-score SME (using means)	248	40	63.719	5.438	.530
Raw consensus (using medians)	248	40	188.692	11.596	.640
Raw SME (using medians)	248	40	187.663	11.396	.580
Self-Report					
College GPA Category	682	1	2.86	.850	--
Archival Data					
LDAC performance Score	178	1	88.961	7.004	--
ROTC National OML Score	178	1	82.496	8.324	--
SAT/ACT	70	1	1126.430	128.323	--
Whole Person Score	85	1	341.706	367.516	--
College GPA	178	1	3.266	0.368	--

Note. College GPA Category 1 = 2.0-2.5, 2 = 2.6-3.0, 3 = 3.1-3.5, 4 = Above 3.5.

^aNo reliability estimate is provided for the officer fit index because the work value scales that comprise it are partially ipsative in nature and using traditional indexes of reliability such as split-half or coefficient alpha would be inappropriate.

Table 4.4. *Predictor Score Subgroup Differences*

Predictor	M-F	WNH-B
Figural Reasoning	.01	.67
Objective-Format Consequences		
Deviation Score	.13	1.02
C-score	.16	1.17
Rational Biodata Inventory (RBI)		
Army Affective Commitment	.08	-.43
Achievement Orientation	-.23	-.17
ARC Hostility	-.06	.58
Career Expectations	.18	-.66
Fitness Motivation – All Items	.75	.22
Fitness Motivation – Old	1.10	.26
Fitness Motivation – New 1	.40	.15
Fitness Motivation – New 2	.48	.22
Goal Orientation	.00	-.60
Tolerance to Injury	.73	.57
Peer Leadership	.02	-.09
Self- Efficacy	-.08	-.27
Tolerance for Ambiguity	-.36	-.56
Writing	-.14	.18
MCPA-Phrases		
Raw Total	.12	-.13
Adj. Raw Total	.14	-.13
Profile Similarity Index	-.02	-.03
Leader Knowledge Test (LKT)		
Characteristics C-score	-.17	.44
Skills C-score	-.02	.46
Characteristics Elevation	-.09	-.59
Skills Elevation	-.17	-.67
Characteristics Scatter	.02	-.18
Skills Scatter	.19	.36

Table 4.4. Predictor Score Subgroup Differences (Continued)

Predictor	M-F	WNH-B
Work Values		
Do work that challenges me	.12	.19
Work in a comfortable, relaxed environment	.28	.15
Do work that keeps me close to home	.19	.25
Provide guidance and direction to others	.07	-.18
Receive a good salary and benefits	.16	-.35
Receive recognition or praise for what I do	.08	-.08
Come up with my own way to do tasks	.07	.38
Contribute to society and the well-being of others	.06	-.18
Have well-defined rules for accomplishing tasks	.12	-.28
Work as part of a team	.39	.02
Work on a variety of types of problems	.46	.06
Profile Similarity Index	.08	-.09
Virtual Leadership Judgment Simulation		
C-score consensus (using medians)	-.22	.13
C-score consensus (using means)	-.18	.24
C-score SME (using medians)	-.17	.15
C-score SME (using means)	-.18	.19
Raw consensus (using medians)	-.01	.47
Raw SME (using medians)	-.02	.47
Self-Report		
College GPA Category	-.22	.26
Archival Data		
LDAC performance Score	.29	.04
ROTC National OML Score	.18	-.16
SAT/ACT	--	--
Whole Person Score	--	--
College GPA	-.05	.09

Note. Sample sizes were: Male (M) $n = 157 - 614$, Female (F) $n = 21 - 69$, White Non-Hispanic (WNH) $n = 127 - 606$, Black (B) $n = 15-58$.

CHAPTER 5: CRITERION-RELATED VALIDATION ANALYSES AND RESULTS

Bethany H. Bynum and Amanda Koch

This chapter examines relationships between the predictor and criterion scores. Sample sizes for the platoon leaders (PLs), company commanders (CCs) and Battalion Staff (at the rank of MAJ) were sufficient to support bivariate validity analyses. In addition, sample sizes of platoon leaders (PLs) and company commanders (CCs) were sufficient to support regression-based analyses. Overall, the four performance dimensions and Career Intentions were well predicted. RBI scales Fitness Motivation, Peer Leadership, Goal Expectations, Hostility to Authority, and Army Identification predicted at least one outcome well. Additionally, LKT, Objective-Format Consequences, and self-report GPA predicted at least one outcome well. The weakest predictors were MCPA-Phrases, work values, and the VLJS (which was administered only to a subset of PLs). To estimate multivariate validities for battalion XO (Bn XO), battalion S3 (Bn S3), and battalion commander (Bn Cdr) positions we constructed synthetic equations. When the same synthetic equations were applied to estimate validity estimates for PL and CC positions, the synthetic validities were similar but slightly lower than those obtained in regression analyses, suggesting that this approach yields accurate validity estimates for the higher echelon positions. The actual synthetic validity estimates for the higher echelon positions indicated that the predictor measures are valid predictors of performance at the higher in these positions.

Analysis Approach

We examined bivariate relationships between predictor and criteria and determined the optimal predictor battery for maximizing the prediction of (a) the four performance factor scores (with method variance removed as described in Chapter 3) and (b) the Career Intentions criterion variable.

Bivariate Correlations

Using the raw data, bivariate correlations were estimated separately for PLs, CCs, and battalion staff (i.e., Bn XOs and Bn S3s combined). Pairwise deletion was used such that all useable data were included for each predictor-criterion relationship. Predictors included those described in Chapter 4, and criteria include the four factor scores and Overall Performance composite described in Chapter 3, archival outcome variables, and Career Intentions.

Tables 5.1, 5.2, and 5.3 provide the predictor-criterion correlations for PLs, CCs, and battalion level officers, respectively. A number of predictors showed significant relationships with Overall Performance for all three sets of positions. The three RBI Fitness Motivation scales (Old, New 1, and New 2) were the strongest predictors of Overall Performance for all positions. Several other RBI scales also predicted Overall Performance significantly in all positions, namely Achievement Orientation, Goal Expectations, Peer Leadership, and Tolerance for Ambiguity. The LKT, which was not administered to Bn XOs and Bn S3s, was significantly related to overall performance for both PLs and CCs.

Table 5.1. *Bivariate Validities for Platoon Leader*

Predictor	TTP	MACP	PFE	LSPD	Overall Performance	Promotion to 1LT	Awards Std. Total	Awards Std. Weighted	Career Intentions
Figural Reasoning	.08	.15	-.02	.09	.06	-.10	-.11	-.07	-.06
Objective Consequences Test	.15	.22	.03	.16	.13	-.05	-.09	-.08	-.09
RBI: Achievement Orientation	.17	.13	.20	.15	.19	-.01	.05	.07	.26
RBI: Hostility to Authority	.01	.05	-.06	-.01	-.01	.05	-.19	-.17	-.25
RBI: Army Identification	-.01	-.03	.05	-.01	.01	.02	.13	.15	.51
RBI: Career Expectations	-.06	-.06	-.06	-.06	-.07	.06	.23	.19	.87
RBI: Fitness Motivation (Old)	.35	.15	.51	.32	.40	-.04	-.17	-.11	.01
RBI: Fitness Motivation (New 1)	.29	.13	.44	.26	.34	-.04	-.13	-.08	.03
RBI: Fitness Motivation (New 2)	.33	.17	.47	.31	.38	-.05	-.15	-.09	.04
RBI: Generalized Self-Efficacy	.08	.04	.11	.07	.09	-.08	.01	.03	.12
RBI: Goal Expectations	.10	.11	.07	.11	.10	.01	.06	.05	.52
RBI: Peer Leadership	.12	.18	.02	.13	.11	.04	.05	.03	.09
RBI: Tolerance for Injury	.19	.10	.21	.16	.19	-.03	-.07	-.03	-.04
RBI: Tolerance for Ambiguity	-.12	-.11	-.08	-.11	-.11	.05	-.04	-.08	-.01
RBI: Written Communication	.04	.12	-.03	.06	.03	-.04	-.02	-.04	.00
MCPA-Phrases	-.02	-.03	.02	-.01	-.01	.08	-.01	.00	.02
LKT: Characteristics	.15	.17	.10	.14	.14	-.06	-.01	.00	-.02
LKT: Skills	.16	.17	.09	.15	.14	.02	.03	.07	-.08
Work Values	.00	-.03	.03	.00	.01	.08	-.07	-.03	.02
College GPA (self-reported)	.11	.14	.07	.11	.11	-.14	.00	-.03	.02
Virtual Leadership Simulation	.04	.00	.07	.01	.04	.00	.04	.07	.03
<i>Archival Data</i>									
National ROTC OML	.28	.26	.25	.26	.28	-.14	.14	.16	-.07
SAT/ACT Score	-.05	.03	-.11	-.03	-.06	--	--	--	-.13
College GPA	.03	.06	.01	.02	.03	-.07	.18	.18	-.13
ROTC LDAC Performance	.16	.19	.08	.15	.15	-.03	.24	.27	.04
ROTC Whole Person Score	-.14	-.16	-.16	-.13	-.16	.15	--	--	.21

Note. Sample $n = 175$ -465; VLJS $n = 81$ -175; Archival data $n = 60$ -175. TTP = Technical Task Proficiency. MACP = Management, Administration and Communication Proficiency. PFE = Physical Fitness and Effort. LSPD = Leadership, Supervision, and Personal Discipline. Samples sizes less than 50 are not reported. Bolded values indicated $p < .05$ (two-tailed).

Table 5.2. Bivariate Validities for Company Commander

Predictor	TTP	MACP	PFE	LSPD	Overall Performance	Promotion to 1LT	Promotion to CPT	Awards Std. Total	Awards Std. Wtd.	MMT	Career Intentions
Objective Consequences Test	.15	.24	.03	.16	.13	-.01	.19	-.08	-.08	.11	-.16
RBI: Achievement Orientation	.23	.20	.20	.23	.23	.03	-.06	.13	.13	-.05	.33
RBI: Hostility to Authority	.00	-.02	.05	-.01	.01	.08	.04	-.02	.01	-.22	-.28
RBI: Army Identification	.01	.00	.01	.01	.01	.10	-.03	.07	.05	-.02	.47
RBI: Career Expectations	-.01	-.04	.00	-.03	-.02	-.04	-.10	.24	.19	.08	.86
RBI: Fitness Motivation (Old)	.49	.28	.65	.47	.55	.04	.03	.12	.12	.05	.06
RBI: Fitness Motivation (New 1)	.44	.24	.60	.42	.50	.03	-.03	.15	.17	.05	.12
RBI: Fitness Motivation (New 2)	.49	.28	.65	.47	.55	.04	.00	.15	.16	.07	.11
RBI: Generalized Self-Efficacy	.21	.17	.19	.20	.21	.12	-.03	.12	.13	-.01	.18
RBI: Goal Expectations	.27	.23	.23	.26	.26	.08	-.04	.12	.09	.04	.41
RBI: Peer Leadership	.24	.24	.16	.23	.23	.11	.04	.12	.09	-.04	.17
RBI: Tolerance for Injury	.14	.06	.20	.12	.16	.01	-.04	.00	.04	.06	-.01
RBI: Tolerance for Ambiguity	-.16	-.19	-.09	-.15	-.14	.10	-.02	-.11	-.09	-.06	-.01
RBI: Written Communication	.12	.24	-.03	.14	.10	-.03	-.02	.00	-.02	.08	.08
MCPA-Phrases	.00	.01	-.02	.00	-.01	.02	.05	.04	.03	-.01	.08
LKT: Characteristics	.21	.27	.09	.21	.19	-.02	.10	-.04	-.04	.14	.04
LKT: Skills	.18	.27	.04	.19	.16	-.02	.15	-.12	-.12	.12	.03
Work Values	.04	.04	.04	.05	.05	-.12	-.16	-.01	.02	.08	.10
College GPA (self-reported)	.20	.23	.11	.21	.19	-.14	-.03	.12	.13	-.01	.11

Note. Sample $n = 115-217$. TTP = Technical Task Proficiency. MACP = Management, Administration and Communication Proficiency. PFE = Physical Fitness and Effort. LSPD = Leadership, Supervision, and Personal Discipline. Bolded values indicated $p < .05$ (two-tailed).

Table 5.3. Bivariate Validities for Battalion Staff Positions

Predictor	TTP	MACP	PFE	LSPD	Overall Performance	Promotion to 1LT	Promotion to CPT	Awards Std. Total	Awards Std. Wtd.	Career Intentions
Figural Reasoning	.04	.09	.05	.08	.08	-.02	.09	-.25	-.22	-.17
Objective Consequences Test	.00	-.02	-.03	-.01	-.02	-.02	.01	-.15	-.14	-.13
RBI: Achievement Orientation	.36	.33	.36	.37	.37	.08	.01	.06	.09	.22
RBI: Hostility to Authority	-.20	-.24	-.11	-.23	-.23	-.10	-.01	-.08	-.12	-.31
RBI: Army Identification	.12	.04	.18	.10	.09	-.01	-.12	.18	.30	.49
RBI: Career Expectations	.13	.03	.20	.09	.09	.03	-.14	.03	.02	.58
RBI: Fitness Motivation (Old)	.39	.20	.64	.35	.35	.07	.04	-.06	-.03	.16
RBI: Fitness Motivation (New 1)	.30	.14	.55	.28	.28	.12	.14	-.05	-.03	.17
RBI: Fitness Motivation (New 2)	.34	.17	.59	.31	.31	.05	.07	-.08	-.07	.12
RBI: Generalized Self-Efficacy	.31	.34	.35	.36	.36	.13	-.06	.02	.10	.30
RBI: Goal Expectations	.33	.32	.39	.37	.37	.04	-.10	.06	.09	.49
RBI: Peer Leadership	.29	.23	.34	.29	.29	.02	-.14	-.04	.00	.04
RBI: Tolerance for Injury	.19	.11	.30	.18	.18	-.15	-.10	-.13	-.05	.03
RBI: Tolerance for Ambiguity	-.27	-.30	-.21	-.29	-.30	-.09	-.06	.07	.00	-.10
RBI: Written Communication	.28	.22	.22	.25	.25	-.08	-.09	-.07	-.05	-.10
MCPA-Phrases	.01	-.02	.06	.01	.00	.01	-.07	.15	.11	.05
College GPA (self-reported)	.10	.06	.14	.09	.09	-.19	-.09	-.06	-.05	-.13

Note. Sample $n = 78-87$. TTP = Technical Task Proficiency. MACP = Management, Administration and Communication Proficiency. PFE = Physical Fitness and Effort. LSPD = Leadership, Supervision, and Personal Discipline. Bolded values indicated $p < .05$ (two-tailed).

RBI Written Communication and RBI Hostility to Authority were significantly related to Overall Performance for battalion staff, but not for PL and CC positions. RBI Tolerance for Injury and College GPA were both significant predictors of Overall Performance in PL and CC positions, but not for battalion staff positions.

Several RBI scales were significantly related to Career Intentions for all three sets of positions—RBI Goal Expectations, Army Identification, Achievement Orientation, Hostility to Authority, Career Expectations, and Generalized Self-efficacy. Additionally, Peer Leadership was related to Career Intentions for PLs and CCs, and the Objective-Format Consequences Test was negatively related to Career Intentions for CCs.

Regression Analyses

To this point, we have reported multiple scores for many of the predictors. Based on the psychometric data reported in Chapter 4 and the predictor correlations and bivariate validities in Tables 5.1 and 5.2, we selected the one score that appeared most psychometrically advantageous for each predictor for inclusion in regression analyses. Predictors were limited to those that maximized sample sizes. The final set of predictors for inclusion is listed in Table 5.4. Criteria included the four performance dimensions, Career Intentions, and overall performance. “Overall performance” was computed as a weighted sum of the performance dimensions based on criterion weights described in Chapter 3.

Table 5.4. Predictor Measures Included in Regression Analyses

Objective Consequences Test (C-Score)
RBI: Achievement Orientation
RBI: Army Identification
RBI: Fitness Motivation (New 2)
RBI: Goal Expectations
RBI: Tolerance for Injury
RBI: Peer Leadership
RBI: Generalized Self-Efficacy
RBI: Tolerance for Ambiguity
RBI: Written Communication Skills
RBI: Hostility to Authority
MCPA-Phrases Profile Similarity Index (PSI)
LKT: Characteristics (C-score)
LKT: Skills (C-score)
Work Values Profile Similarity Index (PSI)
Self-Report GPA
Virtual Leadership Judgment Simulation (PL only)
Figural Reasoning (PL only)

Data Input

The sample sizes differed across predictor measures. Most notably, only about half of the PLs took the Virtual Leadership Judgment Simulation. Analyses have shown that missing data can

lead to low power and downwardly biased estimates of model parameters (Enders & Bandalos, 2001; Muthén, Kaplan, & Hollis, 1987; Roth, 1994; Schafer & Graham, 2002). Application of traditional methods to deal with missing data (e.g., listwise deletion) results in small samples, potentially biased estimates, and limitations on the interpretability of the results.

To deal with the missing data, we opted to estimate a correlation matrix using full information maximum likelihood (FIML) missing data estimation. Analyses have shown that FIML estimation produces parameters that are less biased than those produced by multiple imputation or that result from listwise or pairwise deletion (Enders, 2001). FIML uses all of the available data to estimate the likelihood value of the estimates for each individual. Specifically, when there are missing observations, matrices are produced that incorporate both observed and missing parameter information. The information associated with the complete portion of the vector is used to estimate the likelihood value of the variables with missing data (Enders, 2006). Note that unlike other missing data estimation methods, data are *not* imputed into the database. The missing data are accounted for by including additional information to estimate the correlation matrix. The correlation matrix was used as the input for the regression analyses.

Bayesian Model Averaging

Bayesian Model Averaging (BMA) is a regression-based approach for determining the relative importance of each predictor taking into account the uncertainty associated with identifying a single best regression model (Raftery, 1995). Specifically, with a large number of predictors, as in this project, a set of models may perform similarly but none of the models may have a high likelihood of occurring, which makes it difficult to select any one model as being the best.

Using BMA, a regression model for every possible combination of the predictors is estimated. For each model the probability of being the “best” for the population given the data at hand is estimated. Regression coefficients are estimated by weighting the coefficients for a single model by the model probability and then averaging across all the models. The model averaging process shrinks coefficients for predictors that fail to commonly arise in the most probable best model, thus reducing the possibility that some predictors could appear in the “best” model simply by chance.

We used two methods for evaluating the importance of each predictor to the overall best model. First, we examined the cumulative probability across all the models including the predictor. A cumulative probability of 100% for any given predictor indicates the predictor was included in all plausible (greater than 0% probability) models, and thus is a critical predictor of the focal criteria (Viallefont, Raftery, & Richardson, 2001). Conversely, a cumulative probability of 0% for any given predictor indicates that the predictor was not included in any plausible model and thus is not important to the prediction of the focal criteria.

Second, we used Johnson's (2000) relative weights to evaluate the importance of each predictor based on the model averaged regression equation. We computed a predicted criterion score (\hat{y}) using the average regression coefficients estimated from the BMA model and the raw predictor scores. We then computed the correlation matrix between \hat{y} s and all predictor scores listed in Table 5.2 and ran the Johnson's relative weights procedure. It re-scaled relative weights to a

proportion metric ranging from 0%-100%. After re-scaling, relative weight estimates can be interpreted as the percentage of the criterion variance (R^2) accounted for by each predictor. For example, if the R^2 value is .30 and the relative weight of a predictor is 25%, then that predictor accounts for 25% of the R^2 value of .30.

Next, we estimated the optimized validity coefficient by computing the predictor composite score using the BMA regression equation, then regressing the criteria onto the predictor composite score. All of the predictors were included in the composite regardless of the size of the regression coefficient.

Finally, we compared the BMA regression results to traditional regression analyses. Multiple regression using ordinary least squares (OLS) was used to estimate the regression coefficient and multiple correlation for each comparison. Each of the 18 predictors for PLs and 16 predictors for CCs were included in each OLS model, respectively.

BMA Results

Tables 5.5 and 5.6 present the results of the BMA analyses for PLs and CCs, respectively. The tables include the model averaged regression coefficient for each predictor, the predictor criticality, and Johnson's (2000) relative weight (RW) based on the average regression coefficient. The discussion below is directed specifically to the results of Johnson's relative weights analyses.

Based on the RWs, the Objective-Format Consequences Test, RBI Fitness Motivation, LKT Skills, and self-report GPA were the strongest predictors of the Management, Administration, and Communication Proficiency (MACP) performance dimension. Of the variance in this dimension that can be explained by the full set of predictors, this subset of predictors accounted for 68% and 63% of the explainable variance for PLs and CCs, respectively. For PLs, RBI Peer Leadership also emerged as a strong predictor accounting for 21% of the explainable variance in MACP.

Similar measures emerged as the strongest predictors for Technical Task Proficiency (TTP) and Leadership, Supervision, and Personal Discipline (LSPD) performance dimensions. RBI Fitness Motivation, the Objective-Format Consequences Test, and RBI Tolerance for Injury were the strongest predictors for PLs, collectively accounting for 76% of the explainable variance in TTP and 80% of the explainable variance in LSPD. For CCs, RBI Fitness Motivation, RBI Goal Expectations, and self-report GPA were the strongest predictors, accounting for 72% of the explainable variance TTP and LSPD.

RBI Fitness Motivation and RBI Tolerance for Injury were the strongest predictors of the Physical Fitness and Effort (PFE) performance dimension, collectively accounting for 88% of the explainable variance for PLs and 87% of the explainable variance for CCs. For both PLs and CCs, RBI Army Identification and RBI Goal Expectations emerged as the strongest predictors of Career Intentions, collectively accounting for 77% and 68% of the explainable variance for PLs and CCs, respectively. Additionally, RBI Hostility to Authority (10%) was among the strongest predictors of Career Intentions for CCs.

Table 5.5. Predictor Composite Results for Platoon Leaders

Predictors	MACP			TTP			LSPD		
	\bar{b}	PC	RW	\bar{b}	PC	RW	\bar{b}	PC	RW
Intercept	-1.62			-2.28			-1.99		
Objective Consequences	.36	99%	25.1%	.24	81%	7.8%	.28	90%	13.8%
<i>Achievement Orientation</i>	.00	5%	3.2%	.00	6%	3.7%	.00	6%	3.5%
<i>Army Identification</i>	-.01	15%	0.3%	-.02	21%	0.4%	-.01	18%	0.4%
<i>Fitness Motivation</i>	.18	93%	10.8%	.44	100%	61.2%	.38	100%	59.3%
<i>Goal Expectations</i>	.00	7%	1.8%	.00	5%	1.0%	.00	8%	1.2%
<i>Tolerance for Injury</i>	.00	6%	1.8%	.00	7%	7.2%	.00	5%	7.0%
<i>Peer Leadership</i>	.22	99%	20.8%	.01	11%	1.4%	.02	21%	1.4%
<i>Self-Efficacy</i>	-.13	63%	2.1%	-.01	8%	3.3%	-.02	14%	2.7%
<i>Tolerance for Ambiguity</i>	-.02	18%	0.6%	-.02	15%	0.9%	-.01	11%	0.7%
<i>Written Communication</i>	.00	5%	3.1%	.00	7%	1.2%	.00	5%	0.8%
<i>Hostility to Authority</i>	.00	9%	0.1%	.00	6%	0.0%	.00	5%	0.0%
MCPA-Phrases	-.01	7%	0.1%	-.01	5%	0.1%	.00	5%	0.1%
LKT: Characteristics	.09	20%	4.7%	.15	31%	2.9%	.10	25%	2.2%
LKT: Skills	.46	85%	11.7%	.26	56%	5.3%	.20	49%	3.9%
Work Values Index	.00	5%	0.2%	.00	5%	0.2%	.00	5%	0.1%
GPA	.09	85%	10.8%	.05	51%	2.2%	.03	42%	1.2%
Figural Reasoning	.00	11%	2.3%	.00	5%	0.9%	.00	5%	1.2%
VLJS	.00	27%	0.6%	.00	6%	0.4%	.00	10%	0.3%
	PFE			Career Intentions			Overall Performance		
	\bar{b}	PC	RW	\bar{b}	PC	RW	\bar{b}	PC	RW
Intercept	-2.52			-1.10			-26.65		
Objective Consequences	.01	9%	0.2%	.00	5%	0.8%	1.71	62%	2.4%
<i>Achievement Orientation</i>	.01	9%	3.8%	.00	5%	7.1%	.04	6%	4.1%
<i>Army Identification</i>	-.01	13%	0.6%	.40	100%	36.9%	-.20	21%	0.4%
<i>Fitness Motivation</i>	.68	100%	79.4%	-.01	13%	0.8%	5.46	100%	69.6%
<i>Goal Expectations</i>	-.01	13%	1.1%	.49	100%	39.9%	.00	5%	1.1%
<i>Tolerance for Injury</i>	.00	5%	8.5%	-.23	100%	3.5%	.02	6%	8.0%
<i>Peer Leadership</i>	-.04	29%	0.7%	-.15	90%	2.3%	.02	5%	1.3%
<i>Self-Efficacy</i>	-.01	10%	3.5%	-.01	8%	2.3%	-.11	10%	3.7%
<i>Tolerance for Ambiguity</i>	.00	5%	0.6%	-.01	12%	0.1%	-.09	10%	0.7%
<i>Written Communication</i>	-.06	46%	0.7%	.00	5%	0.5%	-.08	10%	0.8%
<i>Hostility to Authority</i>	.00	6%	0.0%	-.04	42%	4.1%	.00	5%	0.0%
MCPA-Phrases	.00	4%	0.1%	.00	5%	0.1%	-.04	5%	0.1%
LKT: Characteristics	.10	24%	0.2%	.01	8%	0.2%	1.81	35%	2.1%
LKT: Skills	.04	16%	0.3%	-.18	52%	0.6%	2.29	49%	3.4%
Work Values Index	.00	4%	0.1%	.01	7%	0.2%	.00	5%	0.1%
GPA	.02	28%	0.1%	.00	6%	0.1%	.50	53%	1.4%
Figural Reasoning	.00	11%	0.1%	.00	7%	0.1%	.00	5%	0.5%
VLJS	.00	7%	0.1%	.00	21%	0.5%	.00	5%	0.3%

Note. \bar{b} = average probability weighted beta estimate across all possible models. PC = Predictor criticality, the cumulative probability among all models containing the predictor of interest. RW = Johnson's relative importance weight of each predictor based on model averaged regression equation. Italicized predictors are self-report measures administered as part of the Rational Biodata Inventory (RBI). TTP = Technical Task Proficiency. MACP = Management, Administration, and Communication Proficiency. PFE = Physical Fitness and Effort. LSPD = Leadership, Supervision, and Personal Discipline. Sample $n = 466$. Bolded values represent the top three predictors for each performance dimension.

Table 5.6. Predictor Composite Results for Company Commanders

Predictors	MACP			TTP			LSPD		
	\bar{b}	PC	RW	\bar{b}	PC	RW	\bar{b}	PC	RW
Intercept	-2.34			-3.28			-2.96		
Objective Consequences	.36	96%	16.0%	.10	39%	1.7%	.13	52%	2.6%
<i>Achievement Orientation</i>	.00	8%	3.0%	.00	8%	3.5%	.00	8%	3.5%
<i>Army Identification</i>	-.04	30%	0.5%	-.05	34%	0.7%	-.03	26%	0.5%
<i>Fitness Motivation</i>	.25	94%	14.4%	.61	100%	51.4%	.56	100%	50.1%
<i>Goal Expectations</i>	.16	74%	9.3%	.17	75%	11.5%	.14	69%	10.2%
<i>Tolerance for Injury</i>	-.09	49%	1.2%	-.05	31%	4.8%	-.07	43%	4.1%
<i>Peer Leadership</i>	.11	55%	5.8%	.05	29%	3.1%	.05	32%	3.0%
<i>Self-Efficacy</i>	-.01	8%	3.0%	-.01	10%	4.2%	-.01	9%	3.9%
<i>Tolerance for Ambiguity</i>	-.03	24%	2.3%	-.02	19%	0.9%	-.01	14%	0.8%
<i>Written Communication</i>	.01	10%	4.2%	.00	7%	1.2%	.00	7%	1.4%
<i>Hostility to Authority</i>	-.01	12%	0.3%	-.01	10%	0.3%	-.01	11%	0.3%
MCPA-Phrases	.00	7%	0.2%	.01	7%	0.3%	.01	7%	0.3%
LKT: Characteristics	.15	23%	6.9%	.18	28%	3.0%	.12	22%	3.0%
LKT: Skills	.71	86%	13.5%	.44	63%	4.0%	.47	70%	4.9%
Work values Index	.00	7%	0.1%	.00	7%	0.0%	.00	7%	0.0%
GPA	.17	98%	19.4%	.15	93%	9.3%	.15	95%	11.3%
	PFE			Career Intentions			Overall Performance		
	\bar{b}	PC	RW	\bar{b}	PC	RW	\bar{b}	PC	RW
Intercept	-3.49			-1.26			-34.20		
Objective Consequences	.00	7%	0.3%	-.08	40%	1.4%	.43	22%	0.4%
<i>Achievement Orientation</i>	.00	7%	2.4%	.01	9%	8.8%	-.01	8%	2.9%
<i>Army Identification</i>	-.01	10%	0.3%	.33	100%	38.0%	-.32	24%	0.4%
<i>Fitness Motivation</i>	.99	100%	79.1%	.00	8%	0.9%	7.64	100%	65.9%
<i>Goal Expectations</i>	.01	13%	3.8%	.31	100%	29.5%	1.13	57%	7.2%
<i>Tolerance for Injury</i>	-.03	24%	8.0%	-.01	15%	0.4%	-.52	32%	5.7%
<i>Peer Leadership</i>	.00	8%	1.0%	-.01	12%	3.9%	.30	22%	1.8%
<i>Self-Efficacy</i>	-.01	11%	3.8%	-.08	38%	4.0%	-.13	11%	4.1%
<i>Tolerance for Ambiguity</i>	-.01	9%	0.2%	-.01	11%	0.2%	-.19	16%	0.6%
<i>Written Communication</i>	-.02	18%	0.0%	.00	7%	0.5%	-.02	7%	0.8%
<i>Hostility to Authority</i>	.00	7%	0.3%	-.09	69%	10.3%	-.04	8%	0.2%
MCPA-Phrases	.01	7%	0.4%	.00	7%	0.9%	.08	7%	0.3%
LKT: Characteristics	.01	8%	0.2%	-.02	9%	0.3%	1.59	27%	1.8%
LKT: Skills	.01	7%	0.1%	.01	8%	0.0%	2.59	44%	1.5%
Work values Index	.00	6%	0.0%	.00	7%	0.5%	.01	7%	0.0%
GPA	.02	20%	0.1%	.02	28%	0.2%	1.20	82%	6.2%

Note. \bar{b} = average probability-weighted beta estimate across all possible model. PC = Predictor criticality, the cumulative probability among all models containing the predictor of interest. RW = Johnson's relative importance weight of each predictor based on model averaged regression equation. Italicized predictors are self-report measures administered as part of the Rational Biodata Inventory (RBI). TTP = Technical Task Proficiency. MACP = Management, Administration, and Communication Proficiency. PFE = Physical Fitness and Effort. LSPD = Leadership, Supervision, and Personal Discipline. Sample $n = 217$. Bolded values represent the top three predictors for each performance dimension.

For both PLs and CCs, a single RBI scale, Fitness Motivation, was the strongest predictor of overall performance, accounting for 70% and 66% of the explainable variance (PLs and CCs, respectively) accounted for by the entire set of predictors (PL $R^2 = .16$; CC $R^2 = .35$). This is consistent with other research focusing on ROTC and OCS programs, in which RBI Fitness

Motivation is routinely the strongest or one of the strongest predictors of performance (Allen & Young, 2012; Allen et al., 2012; Russell & Tremble, 2011; Putka, 2009). The finding is even more noteworthy in the present effort because it included a much richer set of criterion variables than were available in the other research projects. In fact, the pervasive impact of Fitness Motivation as a predictor across many and varied criterion variables in this project was unexpected. One possible interpretation is that the RBI Fitness Motivation construct measures an element of motivation that underlies every aspect of performance as an officer. Of course, we must acknowledge that there could be issues with our criterion variables, in spite of having a rich set of criterion measures. For example, the performance dimensions are heavily influenced by supervisor ratings, and these were of unknown reliability. (We did not have opportunity to collect data from multiple raters; therefore the reliability of the ratings cannot be estimated.) Furthermore, as noted in Chapter 3, in spite of our efforts to model the performance domain, some of the performance dimensions in the 4-factor model were highly correlated with each other, even after removing method variance. This, of course, leads to comparable patterns of prediction across the criterion variables. We were not able to achieve strong differentiation among the performance factors. The finding that RBI Fitness Motivation is a ubiquitous predictor is one that needs further research.

Table 5.7 provides the optimized validity coefficients based on the model averaged regression equation for each outcome. Overall, the outcomes were well predicted. The predictor set accounted for 12%-41% of the variance in the outcomes for PLs and 30%-34% of the variance in the outcomes for CCs.

Table 5.7. *Validity Evidence of Optimized Predictor Composites*

	BMA <i>R</i>	Full Regression <i>R</i>	Full Regression Adj. <i>R</i>
<i>Platoon Leaders</i>			
MACP	.369	.431	.355
LSPD	.343	.429	.353
TTP	.368	.671	.638
PFE	.487	.526	.471
Overall Performance	.404	.474	.408
Career Intentions	.638	.673	.641
<i>Company Commanders</i>			
MACP	.548	.583	.492
LSPD	.573	.621	.542
TTP	.586	.628	.551
PFE	.635	.680	.617
Overall Performance	.585	.654	.583
Career Intentions	.578	.611	.529

Note. TTP = Technical Task Proficiency. MACP = Management, Administration, and Communication Proficiency. PFE = Physical Fitness and Effort. LSPD = Leadership, Supervision, and Personal Discipline. BMA *R* value is for the predictor composite computed using the model averaged equation predicting the respective criterion dimension. The Full Regression *R* is the multiple correlation of the OLS regression model including all predictors. Adj. *R* estimated population cross-validity for the OLS regression model including all predictors. Sample *n* = 466 for PL and *n* = 217 for CC.

Table 5.6 also includes the multiple correlations (R) for the regression analyses that include all available predictors for each sample (PL or CC). Generally, after correcting for shrinkage, the multiple correlations were similar to those using the model averaged regression equations. Adjusted R s ranged from .35 to .64 for PLs and .49 to .62 for CCs. Tables C.1 and C.2 in the appendix provide the unstandardized and standardized regression coefficients for the PL and CC models, respectively. For many predictors, the unstandardized regression coefficients were similar to averaged coefficients produced using the BMA approach. Differences in the estimates, where the beta estimate is larger for the full model, is likely a result of the model having a low probability of being the best model. Because the BMA coefficients are weighted by the probability of the model itself, if a model has a low probability, then the coefficients will be reduced in size prior to averaging. A single regression model may suggest that a predictor is highly related to the outcome, but that model may have a very low probability of being the best model. Therefore, the BMA estimates reflect a more stable estimate of the relationship between the predictor and criteria for the model. However, the two sets of weighting coefficients (i.e., computed using either the BMA approach or the single regression model) were highly similar. For example, the correlation between the two sets of coefficients for the Company Commanders computed for the six criteria ranged from .73 to .92, and the median correlation was .90.

Summary

Overall, the four performance dimensions, Overall Performance, and Career Intentions were well predicted. RBI scales Fitness Motivation, Peer Leadership, Goal Expectations, Hostility to Authority, and Army Identification predicted at least one outcome well. Additionally, the LKT, Objective-Format Consequences Test and self-report GPA predicted at least one outcome well. The weakest predictors include MCPA-Phrases, Work Values, and the VLJS. Generally, three to four predictors accounted for the most variance in the prediction of an outcome, suggesting that a smaller subset of predictors could be administered and used in the selection of officers without sacrificing validity. There were differences in which predictors emerged as the strongest among the performance dimensions. These results suggest that different attributes are related to different aspects of job performance, as we would expect. That said, one scale in particular, Fitness Motivation, demonstrated the strongest validities for predicting most of the performance dimensions for both PLs and CCs. Still, to identify officers who will be successful on all aspects of their job, the Army will need to consider the profile of traits linked to a number of performance requirements.

There were distinctions and similarities among the best predictors for PLs and CCs. The largest distinction occurred for the TTP and LSPD dimensions, where the profile of predictors best predicting these dimensions varied by position. Interestingly, these two dimensions were also ones judged to differ in importance across the two ranks, with TTP judged to diminish in importance from PL to CC and LSPD judged to increase in importance from PL to CC. The results suggest that performance outcomes are predicted differently at different positions. Accordingly, when choosing or refining officer accession instruments and procedures, the Army must consider how job requirements will change over the course of an officer's career.

Synthetic Equations

Synthetic validation is a strategy used to estimate validities when sample sizes or other factors impede direct estimation of validity. In this project, adequate samples of officers in the three battalion-level positions (XO, S3, and CDR) were not sufficient to support regression analyses; therefore, we implemented a synthetic validation approach to estimate validities for these positions.

In general, synthetic validation involves three main activities: (a) profiling all jobs on the same set of performance dimensions, (b) choosing predictors for each dimension, and (c) building and evaluating the synthetic equations. In Chapter 3, we described our approach for profiling (or weighting) the performance dimensions for the five positions (see Table 3.8). This section describes our approach for choosing predictors and creating and evaluating synthetic equations.

Choosing Predictors for Each Performance Dimension

Following an approach laid out by Johnson (2007), we used (a) bivariate validities and (b) expert judgments about the relationship between each predictor and each performance dimension to identify a predictor battery for each performance dimension. Bivariate validities were computed between each predictor and performance dimension score separately for each position (see Tables 5.1 - 5.3). We then sample size weighted and averaged these correlations across positions to arrive at an overall bivariate correlation between each predictor and performance dimension score.

The experts we provided to each of these raters with descriptions of each predictor, including the instructions to examinees, the scoring approach, the number of items, and sample items to the raters. We asked raters to complete two rating exercises. In the first, they estimated the relationship between each predictor and performance dimension based on the dimension definitions in Table 3.2. That rating was made at the performance dimension, or construct level. The second exercise asked raters to also consider the *type of criterion measure* when making their judgments. It asked raters to estimate the relationship between each predictor and each of three types of measures (experience, performance, or knowledge) for each performance dimension. Experience measures were criterion measures relating to prior work experience. They were typically self-report measures. Performance measures were mostly supervisor ratings, and knowledge measures were the job knowledge tests described in Chapter 3. Ratings for both exercises were made on a 0 to 4 scale ranging from *Not at all Relevant* to *Extremely Relevant*. judgments were made by 11 research psychologists with experience in test development and validation. Raters also indicated if they expected the relationship to be positive or negative. We computed a mean rating for each predictor-performance dimension combination for each exercise.

We used these two sets to decide how to weight the predictors for each performance dimension. of ratings in conjunction with the predictor-performance dimension correlations to We devised a weighting system that gave more weight to predictors with (a) higher bivariate correlations, (b) higher ratings in the expert judgment exercises, and (c) higher ratings in *both* the construct-level exercise and the type of measure exercise. The possible predictor weights ranged from

0 to 7 (see Table 5.8). The weights assigned to each predictor for each performance dimension can be found in Table 5.9. For example, the RBI Fitness Motivation scale received a weight of 7 for the Physical Fitness and Effort (PFE) performance dimension because it had a bivariate correlation greater than .20 with that dimension *and* it achieved a mean rating of 2.5 or greater on both of the expert judgment exercises.

Building and Evaluating Synthetic Equations

Creating synthetic equations for each position was a matter of applying predictor weights and performance dimension weights to create a final set of proportional weights (shown in Table 5.10). Toward that end, we multiplied the predictor weights in Table 5.9 (ranging from 0 to 7) by the performance dimension weights found in Table 3.7 (ranging from 1 to 3) for each position. We then summed across the four performance dimensions for each predictor to arrive at a final weight for each predictor measure for each position. See Table 5.10 for the proportional weight that each predictor received in the predictor battery for each position; relative weights from the BMA analyses for the PL and CC positions are also provided for comparison purposes.

Table 5.8. *Synthetic Predictor Weighting Scheme*

Mean Bivariate Correlation	Mean Expert Judgment Ratings						
	< 1.5	1.5 – 1.9	1.5 – 1.9	2.0 – 2.4	2.0 – 2.4	≥ 2.5	≥ 2.5
	2 exercises	1 exercise	2 exercises	1 exercise	2 exercises	1 exercise	2 exercises
< .10	0	0	0	0	0	1	2
.10 - .14	0	0	0	1	2	3	4
.15 - .19	0	1	2	3	4	5	6
≥ .20	1	2	3	4	5	6	7

Note. The expert judgment rating scale ranged from 0 to 4, where 0 = Not at all Relevant, 1 = Minimally Relevant, 2 = Somewhat Relevant, 3 = Relevant, and 4 = Extremely Relevant.

When comparing the proportional synthetic weights across positions, some differentiation is noticeable. For example, RBI Achievement Orientation and RBI Fitness Motivation receive more weight in the predictor batteries for the more junior positions, and RBI Written Communication Skills and LKT Characteristics and Skills receive more weight in the predictor batteries for the more senior positions.

Comparing synthetic weights to relative weights, we found that some of the same predictors are among the most heavily weighted for both approaches, particularly for PLs. However, there are noticeable differences in the amount of weight some of the predictors receive between the regression-based and synthetic approaches. For example, RBI Fitness Motivation is more heavily weighted in the relative weights approach, whereas RBI Achievement Orientation is more heavily weighted in the synthetic approach.

Table 5.9. *Synthetic Predictor Weights by Performance Dimension*

Predictor	TTP	MACP	PFE	LSPD
Figural Reasoning	2	0	0	0
Objective Consequences	0	1	0	0
RBI: Achievement Orientation	3	1	6	4
RBI: Army Identification	0	0	0	0
RBI: Fitness Motivation	1	1	7	1
RBI: Goal Expectations	0	0	3	1
RBI: Tolerance for Injury	0	0	1	0
RBI: Peer Leadership	0	2	0	5
RBI: Generalized Self-Efficacy	0	0	1	0
RBI: Tolerance for Ambiguity	0	0	0	0
RBI: Written Communication Skills	0	5	0	0
RBI: Hostility to Authority	0	0	0	0
MCPA-Phrases	0	0	0	0
LKT: Characteristics	0	1	0	2
LKT: Skills	0	1	0	2
Work Values Index	0	0	0	0
College GPA (self-reported)	1	0	0	0
Virtual Leadership Judgment Simulation	0	0	0	0

Finally, the overall validity coefficient was calculated for the test battery and the overall performance composite for each position. We performed this step by compositing the predictor-performance dimension correlations from step 4 based on (a) the predictor weights (in Table 5.9) and (b) the performance dimension weights (in Table 3.7). Gulliksen's (1950) equation 6 (p.77) was used to derive an overall validity coefficient that was weighted by both predictor and performance dimension weights. Table 5.11 presents the overall validity coefficients for each position; BMA *R* values are presented for comparison purposes. Validity estimates are similar for the two approaches, particularly for PLs. The only validity estimate for Bn XO's, Bn S3s, and Bn Cdrs comes from the synthetic validity analyses. This estimate of .33 is similar to the estimates for PLs and CCs, suggesting that this predictor battery is valid for the battalion-level positions in addition to the more junior positions.

Table 5.10. Comparison of Synthetic and BMA Weights by Position

Predictor	Position				
	Platoon Leaders		Company Commanders		Battalion XO's, S3s, & Commanders
	Synthetic	RW	Synthetic	RW	Synthetic
Figural Reasoning	4.9%	.5%	3.1%	NA	1.9%
Consequences Test	.8%	2.4%	.8%	.4%	2.8%
RBI: Achievement Orientation	29.3%	4.1%	28.7%	2.9%	22.2%
RBI: Army Identification	.0%	.4%	.0%	.4%	.0%
RBI: Fitness Motivation	22.0%	69.6%	20.9%	65.9%	13.0%
RBI: Goal Expectations	8.9%	1.1%	9.3%	7.2%	5.6%
RBI: Tolerance for Injury	2.4%	8.0%	2.3%	5.7%	.9%
RBI: Peer Leadership	9.8%	1.3%	13.2%	1.8%	19.4%
RBI: Generalized Self-Efficacy	2.4%	3.7%	2.3%	4.1%	.9%
RBI: Tolerance for Ambiguity	.0%	.7%	.0%	.6%	.0%
RBI: Written Communication Skills	4.1%	.8%	3.9%	.8%	13.9%
RBI: Hostility to Authority	.0%	.0%	.0%	.2%	.0%
MCPA-Phrases	.0%	.1%	.0%	.3%	.0%
LKT: Characteristics	4.1%	2.1%	5.4%	1.8%	8.3%
LKT: Skills	4.1%	3.4%	5.4%	1.5%	8.3%
Work Values Index	.0%	.1%	.0%	.0%	.0%
College GPA (self-reported)	7.3%	1.4%	4.7%	6.2%	2.8%
Virtual Leadership Judgment					
Simulation	.0%	.3%	.0%	NA	.0%

Note. Synthetic = percentage of the weight the predictor receives in the test battery based on synthetic validity procedures. RW = percentage of the weight the predictor receives in the test battery based on Johnson's relative importance weights (using model averaged regression equation). Battalion-level positions were combined because their weights for the performance dimensions (shown in Table 3.7) were identical.

Table 5.11. Comparison of Synthetic and BMA Validity Coefficients by Position

	Synthetic <i>R</i>	BMA <i>R</i>
Platoon Leaders	.37	.40
Company Commanders	.36	.59
Battalion XO's, Battalion S3s, & Battalion Commanders	.33	NA

Note. Synthetic *R* value is the overall validity coefficient. BMA *R* value is for the predictor composite computed using the model averaged equation.

Summary

We conducted synthetic validity analyses to obtain a validity coefficient for the battalion staff positions, as their small sample sizes prevented us from conducting traditional empirical analyses. The results of the synthetic validity analyses suggest that the predictor battery is likely to be a valid predictor of performance not only for PLs and CCs but also for officers holding battalion staff positions.

CHAPTER 6: SUMMARY AND CONCLUSIONS

Mark Young, Peter Legree, and Laura Ford

The U.S. Army is undergoing a major and remarkable transition in its personnel structure. The Army's end strength is projected to drop significantly through 2017 (Tan, 2016). These changes will increase the challenges and criticality of selecting those officers who have the most promising potential to lead our future, rapidly changing Army.

This report describes one project in ARI's broader effort to enhance the selection, assignment, and retention of officers. The goal of this broader program is to enhance the Army's ability to identify individuals who will perform well as junior officers, excel at higher leadership levels, and be motivated to pursue a career of long-term service.

The current effort builds upon earlier ARI research that (a) identified the major duty requirements (Paullin et al., 2011) across a range of company and field-grade officer leadership positions (e.g., platoon leader [PL], company commander [CC], battalion staff officers [Bn XO, Bn S3], and battalion commander [Bn Cdr]), and (b) developed both predictor and criterion measures relevant to Army officer performance and continuance (Paullin et al., 2012). Using the knowledge and lessons learned from these earlier efforts, the current effort "Identifying and Validating Selection Tools for Officer Performance and Retention" was conducted to further inform the development and implementation of the Army's officer selection measures and procedures, while enhancing the development and understanding of relevant scientific theory.

Objectives

The primary objective of this validation effort, comprising both concurrent and synthetic methods, was to evaluate a battery of individual difference measures for predicting officer job performance, career intentions, and career success. The results have the potential to inform and improve the assessment of officer candidates for commissioning into the Army. In this context, "career success" is defined as (a) performing effectively in early and mid-career positions, and (b) being viewed as capable of reaching the battalion command level. Several viable predictor measures had already been developed that could be administered in a pre-accession setting with a reasonable probability of predicting officer success. We refined and updated several of these existing measures, for example, to make them amenable to computer administration and scoring. We also developed (a) a new simulation to measure leadership potential (VLJS) and (b) a measure of written communication skills.

As officers move from company grade (e.g., platoon leader, company commander) to field grade leadership positions (e.g., battalion command and staff positions), their numbers greatly diminish, making it difficult to conduct criterion-related validation of new predictor composites. Accordingly, we applied a synthetic validation strategy for evaluating the multivariate value of our predictor measures for these higher-level positions. We believe that such an approach is likely to become increasingly valuable to the Army, as it continues to shrink in size while officer job performance requirements change more rapidly than in the past.

A second objective of this effort was to build a database containing predictor data for a sample of officers who are at various points in their Army career. The database was populated with scores from predictor and performance measures collected during this project (including variables available from archival data sources). The database provides a foundation for tracking the career path of these officers, allowing ARI to evaluate the extent to which the predictor scores are related to both short- and long-term career retention, performance at higher ranks, and career achievements.

Throughout this research, we gave considerable attention to the measurement of performance outcomes, and sought a wide range of performance measures in an effort to ensure breadth to the job performance criterion space. This process was theoretically driven, and we used multiple information sources and approaches (e.g., self-reports, supervisor ratings, objective performance tests) in combination for developing the criterion measures used in the validation. This multidimensional construct-oriented approach to criterion development has been used in ARI's previous officer selection research (e.g., Allen et al., 2012; Allen & Young, 2012; Russell et al., 2010) and follows the general approach used in the Army's Project A (Borman & Smith, 2012; Campbell & Knapp, 2010).

Another recurring theme throughout this work was our effort to maximally leverage ARI's existing officer selection measures while striving to improve and expand upon these measures for potential future operational use.

Summary of Validation Approach and Key Findings

The data collection and analysis steps for the validation effort over the 3-year period are summarized below.

- *Data Collection.* Predictor and criterion data were collected from active duty officers at 11 FORSCOM sites and one TRADOC site. They included PLs, CCs, Bn XO, Bn S3s, and Bn Cdrs. According to their duty position, officers completed different combinations of predictor and criterion instruments. Supervising officers completed performance ratings on their subordinate officers (Bn Cdrs were asked to rate their CCs, Bn XO, and Bn S3; CCs were asked to rate their PLs). Of the 877 officers who participated in the data collections, 836 provided data that were useable for at least some portion of the instruments.
- *Criterion Analyses.* To enhance the breadth of coverage of the officer performance domain, numerous criterion variables were combined to form more comprehensive performance dimensions. Through empirical and rational analyses, the variables were reduced to create a final set of criterion scores: a retention criterion (Career Intent) and four broad performance dimensions: (a) Technical Task Proficiency, (b) Management, Administration, and Communication Proficiency, (c) Physical Fitness and Effort, and (d) Leadership, Supervision, and Personal Discipline. A weighting scheme was also developed to differentiate performance dimensions across the five officer positions examined (PL, CC, Bn XO, Bn S3, and Bn Cdr). The weights were used to create an overall performance composite for each position.

- *Predictor Analyses.* To assess officer skills, abilities, and other characteristics (SAOs), a compilation of predictors was adopted. They included (a) existing measures from operational programs and prior research efforts, (b) newly developed measures, and (c) archival data. The predictor measures were the:
 - Rational Biodata Inventory (RBI) – 11 scales
 - Modified Career Path Appreciation (MCPA) – Phrases
 - Leader Knowledge Test (LKT)
 - Work Values
 - Objective-Format Consequences Test
 - College GPA (self-report)
 - Virtual Leadership Judgment Simulation (VLJS)
 - Archival Variables

Overall, the reliability estimates for scores on the predictor measures were quite strong with the exception of the MCPA.

- *Validation Analyses.* Numerous statistical analyses were utilized to examine the relationships between predictor and criterion scores for the sample groups where sample sizes were sufficiently large (PL and CC).
 - First, *bivariate validities* were estimated across predictors and criterion measures for PLs, CCs and Bn XO/s/ S3s.
 - Next, based on the psychometric properties of the predictors, predictor correlations, and bivariate validities, analysts narrowed the multiple scores per predictor to just one—the score that was most psychometrically sound while maximizing sample size. This final scored predictor set along with the final set of criterion measures (four performance dimensions, plus career intent and overall performance) were input into regression analyses. Regression analyses were implemented to determine the optimal predictor battery for maximizing the prediction of the criterion measures.
 - A regression model averaging process, Bayesian Model Averaging (BMA), was implemented to examine the relative performance of each predictor using criterion weights. Different criterion weights were developed to create an overall performance composite for each position (see Table 3.8). Using the BMA approach, a regression model was estimated for every possible combination of predictors. The results included composites of averaged regression coefficients (\bar{b}), predictor criticality (PC), and Johnson's (2000) relative weights (RW). The

RW values identified the percentage of variance accounted for by each predictor (see Tables 5.5 and 5.6), and hence its importance. For both platoon leader and company commander positions, validity evidence was determined by regressing the respective criterion composite onto the predictor composite to compute an optimized validity coefficient (R) for each outcome. Similar regression weights were obtained using conventional regression procedures.

- Finally, a synthetic validation approach was used to analyze data for all positions, but especially those where sample sizes were too small for regression-based approaches (Bn XO, S3, and Cdr). The synthetic validation approach comprised three main activities: (a) conducting a job analysis to identify criterion dimensions by position, (b) having subject-matter experts identify predictors of each criterion dimension, and (c) building and evaluating synthetic equations for each position. The synthetic equations were created by applying predictor weights and performance dimension weights to yield a final set of proportional weights. For the battalion command and staff positions, the only validity estimates from this effort come from the synthetic validity analyses.

In general, both platoon leader and company commander criteria were well predicted, and validity estimates were higher for the company commanders. The RBI scales (Fitness Motivation, Peer Leadership, Goal Expectations, Hostility to Authority, and Army Identification) predicted at least one outcome. In addition, the LKT, Objective-Format Consequences Test, and GPA predicted at least one outcome well. For platoon leaders, the best predictors of performance across all four performance dimensions (from Table 5.1) were the RBI scales for Fitness Motivation, Achievement Orientation, Peer Leadership, and Tolerance for Injury; the LKT scales for Characteristics and Skills; and the Objective-Format Consequences Test. For company commanders, the strongest predictors across these dimensions were the RBI scales for Fitness Motivation, Goal Expectations, Peer Leadership, Achievement Orientation, and Generalized Self-Efficacy; the LKT scales for Characteristics and Skills; and college GPA. For both positions, the predictive strength of the RBI was seen for the career intent criterion. That is, the RBI scales of Goal Expectations, Army Identification, Achievement Orientation, Hostility to Authority, Generalized Self-Efficacy, and Peer Leadership were significantly related to career intentions. Surprisingly, the Objective-Format Consequences Test scores were *negatively* correlated with career intentions among the company commanders. For Battalion XOs/S3s, the RBI Achievement Orientation, Self-Efficacy, and Written Communication Skills scales predicted all four job performance dimensions.

Implications and Future Directions

The findings from this large 3-year validation effort highlight the importance of non-cognitive attributes in predicting the performance of Army officers who are serving at the company and field-grade leadership levels. These encouraging results are consistent with previous findings (e.g., Allen et al., 2012; Allen & Young, 2012; Russell, 2010) and suggest future directions for improving officer selection.

Several scales from the Rational Biodata Inventory performed well in predicting officer performance across a range of leadership positions. Importantly, results from the revised Fitness

Motivation scale show that it has less potential for adverse impact against females while retaining the validity of the original scale. This is a very important finding, especially given that it was shown to be among the best predictors of officer performance. The revised Fitness Motivation scale should be routinely used in future selection research with officers.

Among the new measures developed and evaluated for this effort, the Objective-Format Consequences test performed especially well in predicting selected criteria. This measure was significantly correlated with several performance dimensions across leadership positions. Unfortunately, the measure also showed the highest subgroup differences (with whites scoring higher than blacks) among all the predictors in the battery. Future research should examine the underlying cause for these subgroup differences and clarify the underlying construct which is being measured by the Objective-Format Consequences test (e.g., to what extent is it capturing verbal reasoning vs. divergent thinking). Alternate scoring keys for reducing potential adverse impact – while maintaining validity – should also be explored. Another limitation of this measure is suggested by its *negative* correlation with career intention among the company commanders. Clarifying its construct validity in future research may shed some light on this unexpected finding.

In the current research, it is noteworthy that for those positions for which both synthetic weights and regression-based weights could be estimated (i.e., platoon leaders and company commanders), both approaches yielded comparable validity estimates. This finding is consistent with past research supporting the accuracy of synthetic validity estimates for military (Peterson, Wise, Arabian, and Hoffman, 2001) and nonmilitary jobs (Johnson et al., 2010). Johnson and his associates have argued that synthetic validation methods have been greatly underutilized by industrial-organizational psychologists, and that this approach offers a number of advantages relative to traditional criterion-related validation strategies. These advantages are becoming increasingly significant for the Army as it continues transitioning through a period of rapid change in its personnel structure and dramatically reduces its end strength over the next few years. The use of synthetic validation should be given serious consideration in any future efforts to develop and assess predictor composites for Army officers.

REFERENCES

- Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT battery. *Military Psychology*, 9, 5-37.
- Allen, M. T., Bynum, B. H., Erk, R. T., Babin, N. E., & Young, M. C. (2012). *Selecting soldiers and civilians into the U.S. Army Officer Candidate School (OCS): Developing empirical selection composites* (Technical Report 1343). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Allen, M. T. & Young, M. C. (Eds.) (2012). *Longitudinal validation of non-cognitive measures for the U.S. Army Officer Candidate School* (Technical Report 1323). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Borman, W. C., & Smith, T. N. (2012). The use of objective measures as criteria in I/O psychology. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 532-542). New York: Oxford University Press.
- Bynum, B. H., & Legree, P. J. (2015). *ROTC longitudinal annual report: 2013* (Research Note 2015-01). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Campbell, J. P. (2012). Behavior, performance, and effectiveness in the 21st century. In S. Kozlowski (Ed.), *The Oxford handbook of organizational psychology* (pp. 159-195). New York: Oxford University Press.
- Campbell, J. P., & Knapp, D. J. (Eds.). (2001). *Exploring the limits in personnel selection and classification*. Mahwah, NJ: Lawrence Erlbaum.
- Campbell, J. P., & Knapp, D. J. (2010). Project A: 12 years of R & D. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 865-886). New York: Routledge Taylor & Francis Group.
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt, W.C. Borman, (Eds.) *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass.
- Christensen, P. R., Merrifield, P. R., & Guilford, J. P. (1953). *Consequences test*. Springville, UT: Charlotte Mackley.
- Department of the Army. (2010). *Enlisted promotions and reductions* (AR 600-8-19). Washington, DC: Author.
- Department of the Army. (2013). *The Army strategic planning guidance 2013*. Washington, DC: Author.
- Enders, C. K. (2006). Analyzing structural equation models with missing data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 313-344). Greenwich, CT: Information Age.

- Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, 61, 713-740.
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8, 430-457.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Hogan, J. A. (1991). Physical abilities. In M. D. Dunnette and L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2, pp. 781-831). Palo Alto, CA: Consulting Psychologists Press.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jacques, E., & Stamp, G. (1990). *Development of stratified systems theory for possible implementation in the U.S. Army* (Research Note 90-74). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35, 1-19.
- Johnson, J. W. (2007). Synthetic validity: A technique of use. In S. M. McPhail (Ed.), *Alternative validation strategies: Developing new and leveraging existing validity evidence* (pp. 122-158). San Francisco: Jossey-Bass.
- Johnson, J. W., Steel, P., Scherbaum C. A., Hoffman, C. C., Jeanneret, P. R., & Foster, J. (2010). Validation is like motor oil: Synthetic is better. *Industrial and Organizational Psychology*, 3, 305-328.
- Legree, P. J., Kilcullen, R., Psotka, J., Putka, D., & Gintner, R. N. (2010). *Scoring situational judgment tests using profile similarity metrics* (Technical Report 1272). Arlington, VA: U.S. Army Research Institute.
- Legree, P.J., Kilcullen, R.N., Putka, D.J., & Wasko, L.E. (2014). Identifying the leaders of tomorrow: Validating predictors of leader performance. *Military Psychology*, 25, 292-309.
- Lewis, P. (1993). *Career path appreciation (CPA) data reduction and analysis* (Technical Report 983). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 42, 431-462.

- Paullin, C., Legree, P. J., Sinclair, A. L., Moriarty, K. O., Campbell, R. C., & Kilcullen, R. N. (2014). Delineating officer performance and its determinants. *Military Psychology*, 26, 259-277.
- Paullin, C., Sinclair, A. L., Moriarty, K. O., Vasilopoulos, N. L., Campbell, R. C., & Russell, T. L. (2011). *Army officer job analysis: Identifying performance requirements to inform officer selection and assignment* (Technical Report 1295). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Peterson, N. G., Wise, L. L., Arabian, J., & Hoffman, R. G. (2001). Synthetic validation and validity generalization: When empirical validation is not possible. In J. P. Campbell, & D. J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 411-451). Mahwah, NJ: Lawrence Erlbaum.
- Putka, D. J. (Ed.). (2009). *Initial development and validation of assessments for predicting disenrollment of four-year scholarship recipients from the Reserve Officer Training Corps (ROTC)* (Study Report 2009-06). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163.
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47, 537-559.
- Russell, T. L., & Tremble, T. R. (Eds.). (2011). *Development and validation of measures for selecting Soldiers for the Officer Candidate School* (Study Note 2011-02). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- Tan, M. (2016, February 13). Army 2017 budget: 4 key points for Soldiers. *Army Times*. Retrieved from <http://www.armytimes.com/story/military/pentagon/2016/02/13/army-2017-budget-4-key-points-soldiers/80232530/>
- Viallefont, V., Raftery, A. E., & Richardson, S. (2001). Variable selection and Bayesian model averaging in case-control studies. *Statistics in Medicine*, 20, 3215-3230.
- Wardzynski, C., Lyle, D. S., & Colarusso, M. J. (2009). *Toward a U.S. Army Officer Corps strategy for success: A proposed human capital model focused upon talent*. Carlisle, PA: The Strategic Studies Institute of the U.S Army War College.
- Zaccaro, S. J., Gilrane, V. L., Robbins, J. M., Bartholomew, L. N., Young, M. C., Kilcullen, R. N., Connelly, S., & Young, W. (2012). *Officer individual differences: Predicting long-term continuance and performance in the U.S. Army* (Technical Report 1324). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

APPENDIX A.
SUPPLEMENTAL CRITERION PSYCHOMETRIC TABLES

Table A.1. Criterion Score Means by Position

Criterion	Platoon Leader			Company Commander			Battalion S3			Battalion XO			Battalion Commander		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Technical Factor	466	-.12	.71	217	.23	.72	48	.29	.77	39	.37	.63	0		
Com./ Management Factor	466	-.13	.64	217	.25	.62	48	.39	.80	39	.55	.58	0		
Effort/ Fitness Factor	466	-.06	.76	217	.11	.80	48	.17	.70	39	.08	.74	0		
Leadership/ Discipline Factor	466	-.14	.67	217	.27	.67	48	.38	.83	39	.49	.65	0		
Promotion to 1LT	364	-.03	.99	212	.05	1.05	47	-.03	1.17	38	.08	.95	39	-.01	.53
Promotion to CPT	22	-1.04	1.16	200	.14	.92	44	-.18	1.16	37	-.16	1.08	39	.24	.55
Std. Total Awards	271	-.68	.67	212	.35	.78	46	1.08	.72	38	1.39	.70	39	.21	.82
Std. Weighted Awards	271	-.70	.62	212	.37	.81	46	1.04	.74	38	1.42	.72	39	.22	.82
Graduate from CCC	1	1.00		130	.98	.12	5	1.00	.00	1	1.00		0		
Recycle from CCC	1	.00		124	.03	.18	5	.00	.00	1	.00		0		
Graduate from BOLC	195	.99	.10	10	1.00	.00	2	1.00	.00	0			0		
Recycle from BOLC	193	.04	.19	10	.00	.00	1	.00		0			0		
MMT C-Score	0			206	.31	.19	0			0			0		
Career Intentions Composite	465	.14	.72	217	.48	.59	48	.85	.40	37	.95	.37	0		
Discipline (Y/N)	464	.10	.30	217	.16	.36	48	.17	.38	38	.13	.34	0		

Note. MMT = Mental Models Test.

Table A.2. Criterion Score Intercorrelations

Criterion	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Technical Factor	1.00														
2. Com./ Management Factor	.89	1.00													
3. Effort/ Fitness Factor	.80	.48	1.00												
4. Leadership/ Discipline Factor	.98	.91	.76	1.00											
5. Promotion to 1LT	.04	.05	.00	.05	1.00										
6. Promotion to CPT	.10	.11	.06	.10	.30	1.00									
7. Std. Total Awards	.20	.24	.09	.25	-.01	-.07	1.00								
8. Std. Weighted Awards	.22	.26	.12	.26	-.03	-.09	.96	1.00							
9. Graduate from CCC	-.05	-.09	.00	-.05	-.02	-.04	.06	.06	1.00						
10. Recycle from CCC	.02	-.03	.09	.02	-.07	-.32	.05	.03	–	1.00					
11. Graduate from BOLC	.00	.03	-.07	.00	-.06	–	.03	.05	–	–	1.00				
12. Recycle from BOLC	.07	.02	.12	.06	-.02	–	.04	.01	–	–	-.53	1.00			
13. MMT C-Score	.11	.12	.06	.12	-.02	-.04	-.09	-.08	-.03	.05	–	–	1.00		
14. Career Intentions Composite	.08	.11	.01	.11	.02	-.09	.39	.35	-.01	.14	.04	-.04	.04	1.00	
15. Discipline (Y/N)	-.05	-.05	-.05	-.07	.08	-.03	.04	.03	.06	.02	.04	.01	-.12	.04	1.00

Note. Correlations are based on the total sample, using pairwise deletion. Sample *n*'s ranged from 131 to 770. Bold = $p < .05$. MMT = Mental Models Test. The MMT was administered only to company commanders.

Table A.3. Criterion Score Subgroup Differences

Criterion	Male <i>n</i>	Female <i>n</i>	Male- Female <i>d</i>	WNH <i>n</i>	Black <i>n</i>	WNH- Black <i>d</i>	Hispanic <i>n</i>	WNH- Hispanic <i>d</i>
Technical Factor	696	74	.30	569	70	.51	47	.20
Com./ Management Factor	696	74	.20	569	70	.58	47	.29
Effort/ Fitness Factor	696	74	.26	569	70	.23	47	.07
Leadership/ Discipline Factor	696	74	.31	569	70	.47	47	.22
Promotion to 1LT	640	60	.06	518	65	-.16	45	.03
Promotion to CPT	312	30	.14	253	36	.04	21	-.16
Std. Total Awards	549	57	-.16	447	62	-.15	39	.23
Std. Weighted Awards	549	57	-.14	447	62	-.09	39	.27
Graduate from CCC	120	17	-.14	96	19	.32	11	-.11
Recycle from CCC	114	17	.20	92	18	.23	11	.22
Graduate from BOLC	186	21	-.11	153	14	-.12	12	-.12
Recycle from BOLC	184	20	.21	152	13	.21	11	.21
MMT C-Score	186	20	-.19	152	20	.22	16	.07
Career Intentions Composite	694	73	.15	568	69	-.61	46	-.24
Discipline (Y/N)	693	74	-.05	567	70	-.14	46	.08

Note. Positive *d*-values indicate higher scores for the first group listed, and negative *d*-values indicate higher scores for the second group listed. MMT = Mental Models Test.

APPENDIX B.

SUPPLEMENTAL PREDICTOR PSYCHOMETRIC TABLES

Table B.1. Predictor Score Means by Position

Predictor	Platoon Leader			Company Commander			Battalion XO			Battalion S3			Battalion Cdr		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Figural Reasoning	448	14.68	3.49	NA			37	14.65	3.34	46	14.35	3.37	NA		
Objective-Format Consequences															
Deviation Score	423	7.00	.64	211	7.11	.62	35	7.00	.53	46	7.11	.63	66	7.31	.49
C-score	429	.43	.33	215	.45	.36	36	.41	.35	47	.46	.30	66	.57	.20
Rational Biodata Inventory (RBI)															
Army Affective Commitment	448	3.69	.68	211	3.80	.60	39	4.06	.55	46	3.95	.55	27	4.07	.45
Achievement Orientation	448	4.05	.49	211	4.10	.49	39	4.26	.41	46	4.13	.42	27	4.28	.40
ARC Hostility	448	2.87	.73	211	2.73	.70	39	2.59	.77	46	2.44	.74	27	2.33	.49
Career Expectations	448	2.95	.99	211	3.40	.91	39	3.98	.64	46	3.94	.66	27	3.75	.42
Fitness Motivation - All Items	448	3.84	.52	211	3.71	.49	39	3.63	.49	46	3.68	.55	27	3.55	.61
Fitness Motivation - Old	464	3.97	.57	218	3.82	.57	39	3.67	.54	48	3.79	.60	27	3.64	.70
Fitness Motivation - New 1	464	3.83	.51	218	3.70	.48	39	3.61	.52	48	3.67	.53	27	3.55	.61
Fitness Motivation - New 2	464	3.95	.54	218	3.81	.53	39	3.65	.54	48	3.78	.56	27	3.66	.67
Goal Orientation	448	3.41	.66	211	3.56	.59	39	3.68	.66	46	3.61	.68	27	3.78	.53
Tolerance to Injury	448	3.92	.60	211	3.86	.57	39	3.94	.51	46	3.97	.45	27	3.84	.62
Peer Leadership	448	3.68	.61	211	3.68	.58	39	3.95	.53	46	3.83	.53	27	3.91	.59
Self-Efficacy	448	4.30	.46	211	4.26	.43	39	4.36	.40	46	4.28	.44	27	4.30	.36
Tolerance for Ambiguity	448	3.24	.50	211	3.14	.55	39	2.95	.45	46	3.01	.53	27	2.80	.43
Writing	448	3.32	.60	211	3.35	.58	39	3.23	.63	46	3.33	.70	27	3.10	.61
MCPA-Phrases															
Raw Total	419	3.61	.58	212	3.53	.55	36	3.37	.57	47	3.54	.49	65	3.69	.78
Adj. Raw Total	419	12.64	2.37	212	12.34	2.29	36	11.68	2.33	47	12.41	2.11	65	13.10	2.98
Profile Similarity Index	419	.03	.14	212	.03	.14	36	.06	.15	47	.02	.15	65	-.08	.30
Leader Knowledge Test (LKT)															
Characteristics C-score	423	.76	.16	212	.80	.14	NA			NA			NA		
Skills C-score	420	.69	.19	212	.71	.16	NA			NA			NA		
Characteristics Elevation	424	7.12	.84	212	6.94	.83	NA			NA			NA		
Skills Elevation	424	7.40	1.03	212	7.30	1.01	NA			NA			NA		
Characteristics Scatter	424	6.02	2.59	212	6.27	2.71	NA			NA			NA		
Skills Scatter	424	4.19	2.63	212	4.10	2.38	NA			NA			NA		

(continued)

Table B.1. *Predictor Score Means by Position (Continued)*

Predictor	Platoon Leader			Company Commander			Battalion XO			Battalion S3			Battalion Cdr		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Work Values															
Do work that challenges me	448	-.03	.97	207	.02	.86	NA			NA			NA		
Work in a comfortable, relaxed environment	448	-.14	1.03	207	-.11	.96	NA			NA			NA		
Do work that keeps me close to home	448	-.11	.95	207	-.19	.95	NA			NA			NA		
Provide guidance and direction to others	448	.07	1.07	207	.00	1.07	NA			NA			NA		
Receive a good salary and benefits	448	.03	1.02	207	.00	1.02	NA			NA			NA		
Receive recognition or praise for what I do	448	-.03	1.00	207	-.14	.99	NA			NA			NA		
Come up with my own way to do tasks	448	-.10	.99	207	.00	.79	NA			NA			NA		
Contribute to society and the well-being of others	448	.65	1.15	207	.48	1.14	NA			NA			NA		
Have well-defined rules for accomplishing tasks	448	-.20	1.01	207	-.28	.99	NA			NA			NA		
Work as part of a team	448	-.05	.95	207	.06	.80	NA			NA			NA		
Work on a variety of types of problems	448	.15	.91	207	.26	.81	NA			NA			NA		
Profile Similarity Index	448	.07	.34	207	.12	.30	NA			NA			NA		
Virtual Leadership Judgment Simulation															
C-score consensus (using medians)	248	64.10	5.29	NA			NA			NA			NA		
C-score consensus (using means)	248	64.54	4.95	NA			NA			NA			NA		
C-score SME (using medians)	248	62.86	5.39	NA			NA			NA			NA		
C-score SME (using means)	248	63.72	5.44	NA			NA			NA			NA		
Raw consensus (using medians)	248	188.69	11.60	NA			NA			NA			NA		
Raw SME (using medians)	248	187.66	11.40	NA			NA			NA			NA		
Self-Report															
College GPA Category	464	2.90	.86	218	2.76	.84	38	2.42	.92	48	2.71	.82	27	2.41	.84
Archival Data															
LDAC performance Score	175	88.94	7.00	NA			NA			NA			NA		
ROTC National OML Score	175	82.59	8.36	NA			NA			NA			NA		
SAT/ACT	69	1129.71	126.27	NA			NA			NA			NA		
Whole Person Score	84	343.50	369.35	NA			NA			NA			NA		
College GPA	175	3.27	.37	NA			NA			NA			NA		

Note. NA means Not Administered. College GPA Category 1 = 2.0-2.5, 2 = 2.6-3.0, 3 = 3.1-3.5, 4 = Above 3.5

Table B.2. *Predictor Score Intercorrelations*

Predictor Score	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Figural Reasoning	1.00												
Objective-Format Consequences													
2. Deviation Score	.20	1.00											
3. C-score	.33	.66	1.00										
Rational Biodata Inventory (RBI)													
4. Army Affective Commitment	-.03	-.11	-.07	1.00									
5. Achievement Orientation	.06	-.12	.01	.45	1.00								
6. ARC Hostility	.00	.07	.06	-.28	-.08	1.00							
7. Career Expectations	-.09	-.06	-.08	.56	.29	-.26	1.00						
8. Fitness Motivation - All Items	.06	-.11	-.01	.15	.36	.03	.00	1.00					
9. Fitness Motivation - Old	.04	-.10	-.02	.14	.31	.00	-.01	.94	1.00				
10. Fitness Motivation - New 1	.05	-.12	-.02	.16	.39	.04	.01	.98	.89	1.00			
11. Fitness Motivation - New 2	.05	-.11	-.01	.16	.37	.01	.01	.94	.93	.94	1.00		
12. Goal Orientation	-.04	-.11	-.03	.45	.50	-.13	.49	.26	.23	.27	.25	1.00	
13. Tolerance to Injury	.00	-.02	.08	.22	.27	.04	-.01	.48	.46	.45	.45	.18	1.00
14. Peer Leadership	.00	-.12	-.04	.22	.51	.06	.12	.25	.21	.26	.23	.48	.23
15. Self-Efficacy	.05	-.18	-.02	.29	.59	-.01	.13	.41	.34	.42	.36	.48	.40
16. Tolerance for Ambiguity	-.06	-.07	-.06	.02	-.03	.09	-.01	-.12	-.14	-.10	-.10	-.02	-.19
17. Writing	.11	.05	.13	.03	.31	.07	.05	.10	.06	.12	.11	.12	.02
MCPA-Phrases													
18. Raw Total	-.01	-.05	-.02	-.03	-.06	.05	-.01	.00	-.01	-.02	.00	-.01	.06
19. Adj. Raw Total	-.01	-.05	-.01	-.03	-.07	.04	.00	-.01	-.01	-.03	-.01	.00	.06
20. Profile Similarity Index	-.04	.00	-.03	.07	.02	-.07	.02	-.02	-.01	-.01	-.01	.04	-.02
Leader Knowledge Test (LKT)													
21. Characteristics C-score	.21	.09	.19	.06	.10	-.08	.06	.05	.03	.05	.05	.05	.02
22. Skills C-score	.17	.11	.21	.07	.04	-.05	.03	-.01	.01	-.01	.01	.01	.05
23. Characteristics Elevation	-.08	-.33	-.22	.20	.22	-.05	.10	.15	.10	.16	.12	.20	.05
24. Skills Elevation	-.09	-.32	-.20	.26	.28	-.07	.17	.13	.08	.16	.10	.27	.03
25. Characteristics Scatter	.05	-.18	.11	.09	.18	-.10	.06	.07	.07	.08	.09	.11	.14
26. Skills Scatter	.06	-.07	.18	-.09	.08	.08	-.07	.03	.06	.04	.07	-.01	.13

(continued)

Table B.2. *Predictor Score Intercorrelations (Continued)*

Predictor Score	1	2	3	4	5	6	7	8	9	10	11	12	13
Work Values													
27. Do work that challenges me	.07	.01	.04	.08	.08	-.01	.08	.03	.01	.05	.01	.08	.01
28. Work in a comfortable, relaxed environment	.01	.05	.06	.03	.03	.00	.03	.01	.01	.01	-.02	.08	.00
29. Do work that keeps me close to home	.06	.04	.04	-.04	.02	.05	-.04	-.01	-.03	-.01	-.04	.01	-.06
30. Provide guidance and direction to others	-.13	-.02	-.06	.07	.01	.03	.01	-.01	-.01	-.01	-.03	.05	-.07
31. Receive a good salary and benefits	-.08	-.07	-.03	.06	-.02	-.01	.04	-.04	-.02	-.01	-.01	.06	-.05
32. Receive recognition or praise for what I do	-.10	-.07	-.09	.07	-.06	.02	-.03	-.03	-.02	-.01	-.03	.05	-.01
33. Come up with my own way to do tasks	.13	.08	.08	.02	.13	.09	.02	.07	.02	.08	.04	.08	.09
34. Contribute to society and the well-being of others	-.08	-.03	-.07	.17	.09	.03	.05	.02	.01	.03	.02	.04	-.04
35. Have well-defined rules for accomplishing tasks	-.08	-.06	-.03	.02	.00	-.01	.03	.01	.03	.03	.02	.02	.00
36. Work as part of a team	.07	-.02	-.03	.08	.10	.03	.06	.10	.09	.10	.09	.08	.15
37. Work on a variety of types of problems	.07	-.01	-.01	.04	.05	-.02	.04	.07	.05	.08	.05	.04	.06
38. Profile Similarity Index	.04	-.03	-.05	.05	.04	.00	.04	.05	.03	.05	.05	.00	.05
Virtual Leadership Judgment Simulation													
39. C-score consensus (using medians)	.15	.08	.14	.12	.05	-.17	.02	.05	.07	.06	.05	.17	.10
40. C-score consensus (using means)	.15	.10	.15	.12	.05	-.16	.01	.04	.06	.05	.04	.17	.09
41. C-score SME (using medians)	.13	.07	.09	.11	.04	-.13	.03	.03	.06	.05	.03	.14	.09
42. C-score SME (using means)	.14	.08	.11	.13	.04	-.15	.03	.04	.06	.06	.03	.17	.08
43. Raw consensus (using medians)	.12	.25	.12	.01	-.14	-.05	-.05	-.10	-.08	-.10	-.11	-.05	.00
44. Raw SME (using medians)	.11	.26	.10	-.01	-.14	-.02	-.06	-.12	-.10	-.12	-.15	-.07	-.02
Self-Report													
45. College GPA Category	.09	.07	.03	-.02	.13	-.05	-.01	-.01	.00	.02	.03	-.05	-.01
Archival Data ^a													
46. LDAC performance Score	.07	.03	-.04	-.02	.17	.03	.05	.15	.17	.13	.15	.04	.10
47. ROTC National OML Score	.05	.05	-.01	.02	.22	.01	-.08	.20	.23	.17	.24	.01	.06
48. SAT/ACT	.29	.19	.40	-.25	.08	.03	-.27	-.19	-.19	-.18	-.11	-.05	-.02
49. Whole Person Score	-.16	.02	-.22	-.13	-.03	.11	.14	-.15	-.19	-.15	-.15	.02	-.08
50. College GPA	-.06	.11	-.03	-.08	.05	.03	-.14	-.11	-.10	-.11	-.08	-.09	-.22

(continued)

Table B.2. *Predictor Score Intercorrelations (Continued)*

Predictor Score	14	15	16	17	18	19	20	21	22	23	24	25	26
14. Peer Leadership	1.00												
15. Self- Efficacy	.57	1.00											
16. Tolerance for Ambiguity	-.13	-.09	1.00										
17. Writing	.32	.20	-.07	1.00									
MCPA-Phrases													
18. Raw Total	.00	-.02	.04	-.02	1.00								
19. Adj. Raw Total	.00	-.02	.04	-.02	.99	1.00							
20. Profile Similarity Index	.00	-.02	.03	-.06	-.66	-.65	1.00						
Leader Knowledge Test (LKT)													
21. Characteristics C-score	.03	.09	-.04	.11	-.08	-.07	.03	1.00					
22. Skills C-score	-.05	.04	.00	.08	-.03	-.03	.01	.62	1.00				
23. Characteristics Elevation	.23	.23	.13	.00	.01	.00	.06	.00	.02	1.00			
24. Skills Elevation	.27	.27	.11	.02	-.01	-.02	.06	.05	-.03	.76	1.00		
25. Characteristics Scatter	.08	.21	-.06	.14	-.01	-.01	-.03	.36	.28	-.23	-.05	1.00	
26. Skills Scatter	.03	.08	-.04	.12	-.02	-.02	-.03	.19	.26	-.25	-.38	.61	1.00
Work Values Inventory (WVI)													
27. Do work that challenges me	.05	.06	-.02	.06	-.04	-.04	-.01	.08	.02	.06	.07	-.03	-.06
28. Work in a comfortable, relaxed environment	.02	.02	-.01	.08	-.02	-.02	.02	.04	.08	.03	.05	.03	.01
29. Do work that keeps me close to home	-.01	-.02	.04	.05	-.02	-.03	-.03	-.01	.01	.05	.03	-.08	-.07
30. Provide guidance and direction to others	.04	-.06	.08	.00	-.04	-.04	.01	-.02	-.05	.08	.07	-.01	-.04
31. Receive a good salary and benefits	.04	-.02	.01	.01	-.01	-.02	.01	.01	-.01	.14	.13	-.05	-.09
32. Receive recognition or praise for what I do	.06	-.05	.09	-.02	-.02	-.01	.09	-.04	-.05	.08	.06	-.07	-.05
33. Come up with my own way to do tasks	.10	.07	-.03	.09	.00	.00	-.05	.08	.05	.01	-.01	.01	.01
34. Contribute to society and the well-being of others	.06	-.02	.07	.14	.00	.00	.00	-.02	-.04	.09	.12	.01	-.08
35. Have well-defined rules for accomplishing tasks	.03	-.03	.01	.04	-.01	-.01	.03	-.02	.00	.10	.06	-.03	-.04
36. Work as part of a team	.12	.09	-.01	.07	.00	.00	-.06	.06	.04	.07	.05	.05	.03
37. Work on a variety of types of problems	.10	.09	-.10	.06	.02	.02	-.05	.04	.01	.09	.10	.03	-.03
38. Profile Similarity Index	.06	.07	-.04	-.02	.03	.03	-.08	.04	-.03	.02	.03	.07	.02
Virtual Leadership Judgment Simulation													
39. C-score consensus (using medians)	.10	.13	.00	-.08	-.06	-.04	.01	.29	.15	-.05	.06	.12	-.11
40. C-score consensus (using means)	.08	.10	.00	-.08	-.07	-.06	.01	.29	.16	-.06	.04	.10	-.12
41. C-score SME (using medians)	.09	.09	.01	-.09	-.09	-.07	.04	.30	.16	-.01	.08	.06	-.14

(continued)

Table B.2. *Predictor Score Intercorrelations (Continued)*

Predictor Score	14	15	16	17	18	19	20	21	22	23	24	25	26
Virtual Leadership Judgment Simulation (cont'd)													
42. C-score SME (using means)	.10	.08	.03	-.08	-.09	-.07	.05	.31	.17	-.01	.08	.08	-.13
43. Raw consensus (using medians)	-.05	-.11	-.03	-.12	-.11	-.10	.04	.29	.09	-.13	.00	-.13	-.26
44. Raw SME (using medians)	-.06	-.13	-.02	-.12	-.13	-.12	.05	.29	.10	-.13	-.02	-.17	-.26
Self-Report													
45. College GPA Category	.00	-.02	-.04	.22	-.05	-.05	.01	.02	-.03	.01	-.01	-.04	-.04
Archival Data ^a													
46. LDAC performance Score	.19	.20	.06	.17	-.03	-.02	.11	.03	-.06	.03	-.06	-.04	.01
47. ROTC National OML Score	.07	.13	.04	.15	-.05	-.06	.06	-.03	.01	-.05	-.11	.03	.05
48. SAT/ACT	-.04	.14	-.06	.28	-.06	-.04	.13	.23	.24	-.29	-.42	.16	.32
49. Whole Person Score	-.01	.02	.20	-.03	.10	.10	-.13	-.12	-.23	.00	-.02	-.16	-.01
50. College GPA	-.07	-.15	.08	.09	-.14	-.14	.15	-.07	.04	-.10	-.12	-.08	-.03

(continued)

Table B.2. *Predictor Score Intercorrelations (Continued)*

Predictor Score	27	28	29	30	31	32	33	34	35	36	37	38	39
Work Values Inventory (WVI)													
27. Do work that challenges me	1.00												
28. Work in a comfortable, relaxed environment	.22	1.00											
29. Do work that keeps me close to home	.38	.29	1.00										
30. Provide guidance and direction to others	.13	.09	.12	1.00									
31. Receive a good salary and benefits	.16	.20	.19	.20	1.00								
32. Receive recognition or praise for what I do	.17	.15	.10	.33	.31	1.00							
33. Come up with my own way to do tasks	.43	.21	.26	.16	.12	.18	1.00						
34. Contribute to society and the well-being of others	.11	.08	.07	.24	.09	.14	.21	1.00					
35. Have well-defined rules for accomplishing tasks	.17	.33	.22	.16	.35	.22	.20	.16	1.00				
36. Work as part of a team	.29	.22	.17	.22	.16	.16	.40	.21	.21	1.00			
37. Work on a variety of types of problems	.27	.22	.21	.24	.21	.13	.32	.18	.19	.45	1.00		
38. Profile Similarity Index	.14	-.43	-.41	.24	-.17	-.21	.12	.20	-.30	.41	.43	1.00	
Virtual Leadership Judgment Simulation													
39. C-score consensus (using medians)	.04	.03	.02	.01	-.09	-.10	.05	-.01	-.01	.05	.15	.09	1.00
40. C-score consensus (using means)	.06	.04	.03	.02	-.08	-.09	.07	-.02	-.02	.06	.14	.08	.99
41. C-score SME (using medians)	.06	.04	.06	.02	-.09	-.05	.07	.00	-.03	.05	.12	.06	.92
42. C-score SME (using means)	.06	.06	.06	.03	-.09	-.04	.07	.00	-.01	.05	.13	.05	.94
43. Raw consensus (using medians)	.10	.06	.11	.04	-.10	-.07	.07	-.01	.05	.03	.07	-.01	.70
44. Raw SME (using medians)	.10	.06	.12	.05	-.09	-.05	.09	-.03	.02	.04	.07	.00	.64
Self-Report													
45. College GPA Category	.08	.06	.04	-.01	.05	-.03	.05	.05	.03	-.03	-.07	-.09	-.08
Archival Data^a													
46. LDAC performance Score	.02	-.09	.01	-.06	-.03	.01	-.01	-.03	-.12	.05	-.05	.03	.02
47. ROTC National OML Score	.16	-.01	.09	.03	.18	.09	.02	.04	-.02	.01	-.04	-.06	-.05
48. SAT/ACT	.02	.11	.04	-.19	-.13	-.32	-.09	-.10	-.11	-.16	-.22	-.15	-.04
49. Whole Person Score	.01	.00	.05	.09	-.04	.13	.16	.03	-.01	-.02	.06	-.02	-.04
50. College GPA	.17	.02	.09	.12	.17	.14	.05	.15	-.07	-.01	-.07	-.04	-.10

(continued)

Table B.2. *Predictor Score Intercorrelations (Continued)*

Predictor Score	40	41	42	43	44	45	46	47	48	49	50
Virtual Leadership Judgment Simulation (cont'd)											
40. C-score consensus (using means)	1.00										
41. C-score SME (using medians)	.92	1.00									
42. C-score SME (using means)	.94	.99	1.00								
43. Raw consensus (using medians)	.71	.68	.68	1.00							
44. Raw SME (using medians)	.67	.72	.71	.95	1.00						
Self-Report											
45. College GPA Category	-.05	-.02	-.03	-.03	.02	1.00					
Archival Data ^a											
46. LDAC performance Score	.00	.02	.00	-.03	-.04	.15	1.00				
47. ROTC National OML Score	-.04	.01	-.03	-.11	-.07	.71	.51	1.00			
48. SAT/ACT	-.01	-.19	-.18	-.01	-.15	.06	.18	.09	1.00		
49. Whole Person Score	-.04	.01	-.01	.01	.08	.14	.18	.15	-.06	1.00	
50. College GPA	-.07	-.05	-.07	-.03	.00	.77	.09	.68	.12	.13	1.00

Note. With exception of the Virtual Leadership Judgment Simulation correlations, correlations are based on the total sample of platoon leaders and company commanders; sample n ranged from 449 to 682. Correlations for the VLJS are based on the sample of 248 platoon leaders who took that test. Bolded correlations are significant $p < .05$, two-tailed.

^a Archival data include platoon leaders only and sample sizes range from 44-168.

APPENDIX C.
SUPPLEMENTAL VALIDITY TABLES

Table C.1. Full Validation model Regression Coefficients for Platoon Leaders

	TTP		MACP		PFE		LSPD		Career Intentions		Overall Performance	
	b	β	b	β	b	β	b	β	b	β	b	β
Intercept	-1.68		-.95		-2.35		-1.25		-.04		-20.09	
Figural Reasoning	.00	-.01	.01	.06	-.02	-.09	.00	.00	.00	-.02	-.07	-.03
Objective Consequences	.25	.12	.30	.15	.10	.04	.26	.13	-.02	-.01	2.40	.10
<i>Achievement Orientation</i>	.13	.09	.05	.04	.22	.14	.08	.06	.05	.03	1.65	.11
<i>Army Identification</i>	-.12	-.12	-.09	-.09	-.11	-.09	-.12	-.12	.39	.36	-1.28	-.11
<i>Fitness Motivation</i>	.41	.32	.17	.14	.68	.48	.37	.30	-.10	-.08	5.30	.38
<i>Goal Expectations</i>	.04	.04	.10	.10	-.05	-.04	.08	.08	.51	.47	.30	.03
<i>Tolerance for Injury</i>	.08	.07	.05	.05	.03	.02	.06	.05	-.19	-.15	.63	.05
<i>Peer leadership</i>	.11	.10	.21	.20	-.08	-.06	.13	.12	-.17	-.14	.74	.06
<i>Self-Efficacy</i>	-.21	-.14	-.26	-.19	-.11	-.07	-.23	-.16	-.05	-.03	-2.14	-.13
<i>Tolerance for Ambiguity</i>	-.09	-.06	-.09	-.07	-.05	-.03	-.07	-.05	-.06	-.04	-.81	-.05
<i>Written Communication</i>	-.11	-.09	-.02	-.02	-.15	-.12	-.08	-.07	.01	.01	-1.22	-.10
<i>Hostility to Authority</i>	.02	.02	.04	.04	-.03	-.02	.00	.00	-.08	-.08	.03	.00
MCPA-Phrases	-.15	-.03	-.18	-.04	.01	.00	-.11	-.02	-.02	.00	-1.07	-.02
LKT: Characteristics	.35	.08	.35	.09	.26	.05	.30	.07	.35	.08	3.54	.08
LKT: Skills	.41	.11	.44	.13	.19	.05	.39	.11	-.43	-.12	3.84	.10
Work values Index	-.01	-.01	-.02	-.01	.02	.01	.00	.00	.08	.04	.01	.00
GPA	.09	.11	.09	.13	.07	.08	.08	.11	.02	.02	.95	.11
VLJS	-.01	-.04	-.01	-.10	.01	.03	-.01	-.07	-.01	-.08	-.04	-.03
<i>R</i>	.67		.43		.53		.43		.67		.47	

Note. Sample $n = 466$. b = unstandardized regression coefficient. β = standardized regression coefficient. R = multiple correlation of the full regression model. Italicized predictors are self-report measures administered as part of the Rational Biodata Inventory (RBI). TTP = Technical Task Proficiency. MACP = Management, Administration, and Communication Proficiency. PFE = Physical Fitness and Effort. LSPD = Leadership, Supervision, and Personal Discipline $n = 217$. Bolded values indicate statistical significance at $p < .05$.

Table C.2. Full Validation model Regression Coefficients for Company Commanders

	TTP		MACP		PFE		LSPD		Career Intentions		Overall Performance	
	b	β	b	β	b	β	b	β	b	β	b	β
Intercept	-2.45		-1.68		-2.83		-2.23		-.76		-27.01	
Objective Consequences	.30	.15	.39	.23	.04	.02	.31	.17	-.18	-.11	2.47	.12
<i>Achievement Orientation</i>	.05	.03	.02	.02	.05	.03	.04	.03	.11	.10	.46	.03
<i>Army Identification</i>	-.19	-.16	-.17	-.16	-.11	-.08	-.18	-.16	.32	.33	-1.72	-.14
<i>Fitness Motivation</i>	.65	.48	.29	.25	1.02	.68	.59	.47	.08	.07	7.96	.56
<i>Goal Expectations</i>	.23	.19	.21	.20	.15	.11	.20	.18	.36	.36	2.15	.17
<i>Tolerance for Injury</i>	-.16	-.13	-.15	-.14	-.14	-.10	-.17	-.14	-.09	-.08	-1.73	-.13
<i>Peer leadership</i>	.23	.19	.20	.19	.14	.10	.20	.18	-.07	-.07	2.07	.16
<i>Self-Efficacy</i>	-.17	-.10	-.10	-.07	-.21	-.12	-.13	-.09	-.20	-.15	-1.84	-.11
<i>Tolerance for Ambiguity</i>	-.09	-.07	-.10	-.09	-.05	-.04	-.07	-.05	-.08	-.07	-.79	-.06
<i>Written Communication</i>	-.08	-.06	.04	.04	-.19	-.14	-.04	-.04	.01	.01	-1.01	-.08
<i>Hostility to Authority</i>	-.08	-.08	-.08	-.09	-.02	-.02	-.08	-.09	-.11	-.13	-.66	-.06
MCPA-Phrases	.12	.02	.14	.03	.06	.01	.18	.04	-.03	-.01	1.30	.02
LKT: Characteristics	.25	.05	.32	.07	.11	.02	.16	.03	-.44	-.10	2.07	.04
LKT: Skills	.49	.11	.59	.15	.11	.02	.51	.12	.30	.08	4.13	.09
Work values Index	.07	.03	.07	.03	.06	.02	.10	.04	.04	.02	.82	.03
GPA	.17	.20	.15	.21	.11	.12	.16	.20	.06	.09	1.60	.18
<i>R</i>	.63		.58		.68		.62		.61		.65	

Note. Sample $n = 466$. b = unstandardized regression coefficient. β = standardized regression coefficient. R = multiple correlation of the full regression model. Italicized predictors are self-report measures administered as part of the Rational Biodata Inventory (RBI). TTP = Technical Task Proficiency. MACP = Management, Administration, and Communication Proficiency. PFE = Physical Fitness and Effort. LSPD = Leadership, Supervision, and Personal Discipline $n = 217$. Bolded values indicate statistical significance at $p < .05$.